

Exploring the Functional Relevance of Polymorphisms within the CD14 and IRF-1 Gene for Promoter Activity by Haplotype-Specific Chromatin Immunoprecipitation (HaploChIP)



Dissertation
zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades
„Doctor rerum naturalium“
der Georg-August-Universität zu Göttingen

vorgelegt von
Jasmin Mertens
aus München

Göttingen 2010

D7

Referent:

Prof. Dr. med. Gerhard Hunsmann

Korreferent:

PD Dr. rer. nat. Michael Hoppert

Anleiterin:

Prof. Dr. rer. nat. Sabine Mihm

Tag der mündlichen Prüfung:

Table of Contents	I
Abbreviations	III
 1. Introduction	 1
1.1. Innate immune system	1
1.1.1. PRRs	1
1.1.2. Toll-like receptors (TLRs)	2
1.1.3. TLR expression in different cell types	3
1.1.4. TLR signalling pathways	4
1.1.5. TLR4-MD2-CD14 complex and LPS	5
1.1.6. CD14	6
1.1.7. Interferon regulatory factor 1 (IRF-1)	6
1.2. RNA polymerase II (RNAPII) transcription cycle	7
1.3. Single nucleotide polymorphisms (SNPs)	9
1.3.1. Methods for investigation of rSNPs	10
1.3.2. Association of the polymorphisms CD14 rs2569190 and IRF-1 rs2549009 with various disease conditions	11
1.4. Aims of this work	12
 2. Materials & Methods	 13
2.1. Blood donors	13
2.2. Genotyping	13
2.2.1. Discrimination of variant positions using 5'-nuclease assays	14
2.2.2. Discrimination of variant positions using ARMS-PCR	14
2.2.3. Discrimination of variant positions using allele-specific primers	15
2.3. HaploChIP	17
2.3.1. Cross-linking of proteins to the DNA and enzymatic shearing of chromatin	17
2.3.2. Immunoprecipitation	17
2.3.3. Restriction fragment length polymorphism (RFLP) analysis	18
2.3.3.1. Amplification of purified chromatin fragments	18
2.3.3.2. Restriction endonuclease digestion of amplicons	19
2.3.3.3. Quantification of DNA fragments with a bioanalyzer	22
2.4. Sequencing of the CD14 gene	25

2.5. ASTQ	25
2.6. Haplotyping	25
2.6.1. Haplotyping of polymorphisms within the CD14 gene	25
2.6.2. Haplotyping of polymorphisms within the IRF-1 gene	26
2.7. Statistical analysis	26
 3. Results & Discussion	 27
3.1. Establishing and applying HaploChIP to assess the transcriptional activity at CD14 rs2569190 promoter variants in a natural genomic context	27
3.1.1. CD14 rs2569190 RFLP analyses with low cycle number PCR	29
3.1.2. CD14 rs2569190 RFLP analyses with standard cycle number PCR	32
3.1.3. Additional control – immunoprecipitation by a mock isotype (IgG) antibody	34
3.1.4. Lack of correlation between CD14 rs2569190 promoter activity and allelic CD14 transcription in PBMC	36
3.1.5. Validation of HaploChIP and ASTQ results using the imprinted gene SNRPN	37
3.1.6. Sequencing of the CD14 gene and flanking 5' and 3' prime region	39
3.1.7. HaploChIP analysis with antibodies specific for ser5-P/ser2-P RNAPII - CD14 rs2569190 allele-specific initiation of transcription but absence of allelic elongation in PBMC	40
3.2. Establishing and applying HaploChIP to assess the transcriptional activity at IRF-1 rs2549009 promoter variants in a natural genomic context	43
3.2.1. IRF-1 rs2549009 RFLP analyses with standard cycle number PCR	44
3.2.2. Correlation between allele-specific IRF-1 rs2549009 promoter activity and allelic transcription of IRF-1 in PBMC	47
3.2.3. Further functional studies on IRF-1 rs2549009 gene expression	49
3.2.4. LD between IRF-1 rs2549009 and neighbouring promoter polymorphisms	49
 4. Summary	 52
5. Supplementary material	53
 6. Publications	 57
6.1. Functional impact of endotoxin receptor CD14 polymorphisms on transcriptional activity	57

6.2. Functional relevance of the IRF-1 promoter polymorphism rs2549009 on transcriptional activity in a native genomic environment	68
6.3. TLR3 gene polymorphisms and liver disease manifestations in chronic hepatitis C	77
6.4. Congress contributions	79
 7. References	 80
 Acknowledgements	 85
Curriculum vitae	86

Abbreviation

ARMS	allele-specific tetra-primer amplification refractory mutation system
ASTQ	allele-specific transcript quantification
CD14	cluster of differentiation protein 14
cDNA	complementary DNA
CDK	cyclin dependent kinase
CpG-DNA	cytosine-phosphate-guanine-DNA
CTD	carboxyl-terminal domain
DC	dendritic cell
dsRNA	double-stranded
EMSA	electrophoretic mobility shift assay
gDNA	genomic DNA
GPI	glycosyl phosphatidylinositol
HaploChIP	haplotype-specific chromatin immunoprecipitation
HCV	hepatitis C virus
HWE	Hardy-Weinberg equilibrium
IFN	interferon
IKK	I κ B kinase
IL	interleukin
IRAK	interleukin 1 receptor associated kinase
IRF	interferon regulatory factor
LD	linkage disequilibrium
LPS	lipopolysaccharide
LBP	LPS-binding protein
MAF	minor allele frequency
Med	mediator complex
mCD14	membrane CD14
MD2	myeloid differentiation protein 2
mRNA	messenger RNA
MyD88	myeloid differentiation factor 88
NF	nuclear factor
NK	natural killer
PAMP	pathogen-associated molecular pattern
PBMC	peripheral blood mononuclear cell

PCR	polymerase chain reaction
pDC	plasmacytoid DC
Pro	proline
PRR	pathogen recognition receptor
RNAPII	RNA polymerase II
RFLP	restriction fragment length polymorphism
rSNP	regulatory SNP
sCD14	soluble CD14
Ser	serine
ser2-P/ser5-P	serine 2 / serine 5 phosphorylated
SNP	single nucleotide polymorphism
SNRPN	small nuclear ribonucleoprotein polypeptide N
ssRNA	single-stranded RNA
TFII	transcription factor II
Thr	threonine
TIR	Toll/interleukin 1 receptor
TIRAP	TIR domain containing adaptor protein
TLR	Toll-like receptor
TNF	tumor necrosis factor
TRAM	TIR domain containing adaptor molecule
TRIF	TIR domain containing adaptor inducing IFN- β
Tyr	tyrosine

1. Introduction

1.1. Innate immune system

The human host defence serves two main functions, (i) to generate immune response to invading pathogens and (ii) to suppress development of tumor cells (Tamura et al., 2008). The immune system has been traditionally divided into the innate and the adaptive part. While the adaptive component is organized in two specialized cell classes, T cells and B cells, that provide a very large and extremely diverse repertoire of antigen receptors, the effector mechanisms of the innate part of host immunity are activated immediately after infection and rapidly control the replication of the infecting pathogen (Medzhitov and Janeway, 2000). Recognition of invading pathogens is dependent on a limited number of pathogen recognition receptors (PRRs) that recognize pathogen-associated molecular patterns (PAMPs) such as lipopolysaccharide (LPS) or viral nucleic acids (1.1.1). Signal transduction activated by these PRRs induces various target genes including type I interferons (IFNs), proinflammatory cytokines and chemokines. To carry out this task of cellular response a genetic regulatory network controlled by a specific set of transcription factors dependent on cell type and/or cellular stimuli is provided by the innate immune system (Akira et al., 2006; Szabo et al., 2006; Tamura et al., 2008). The functional diversity of transcription factors is dependent on its modification and/or interaction with other transcription factors activated in the same cell type and such regulatory networks are critical for host defence against extracellular pathogens as they rapidly alter the expression of relevant genes (Taniguchi et al., 2001).

1.1.1. PRRs

Invading pathogens are detected by PRRs, which recognize highly conserved structures only produced by microorganisms, like LPS, peptidoglycan, lipoteichoic acids, double-stranded RNA (dsRNA) and glucans. The PRRs are germ-line-encoded receptors which mean that their specificity for the classes of infectious pathogens is genetically predetermined. The expression of these receptors is not clonal. This fact accounts for the rapid kinetics of the innate immune response, thus once an effector cell identifies PAMPs the effector function can be performed immediately and not after proliferation of the acting cell (Medzhitov and Janeway, 2000; Szabo et al., 2006).

PRRs can be classified by structure or by function. Structurally, these receptors belong to several protein families, such as leucine-rich repeat domains, calcium-dependent lectin domains and scavenger-receptor protein domains. Functionally, PRRs can be divided into three classes: secreted, endocytic and signalling (Medzhitov and Janeway, 2000).

On recognizing the pathogens, secreted PRRs flag the microbial cell for the complement system and phagocytes, endocytic PRRs mediate the uptake and delivery of the pathogens into lysosomes and signalling receptors activate signal-transduction pathways that induce the expression of a variety of immune-response genes (Medzhitov and Janeway, 2000).

1.1.2. Toll-like receptors (TLRs)

The Toll-like receptor (TLR) family is the best characterized class of PRRs in mammals and the number of TLR genes ranges between 10 and 15 in most mammalian species (Iwasaki and Medzhitov, 2004). TLRs are widely expressed on immune cells and recognition of PAMPs induces a prompt and potent innate immune response signalled through adaptor molecules, like myeloid differentiation factor 88 (MyD88), to activate transcription factors like nuclear factor (NF)- κ B or interferon regulatory factors (IRFs) that induce expression of antibacterial or antiviral response genes.

TLR1, TLR2, TLR4-6 are expressed on the cell surface and TLR3, TLR7-9 are located on the endosome-lysosome membrane. TLR1, TLR2 and TLR6 recognize bacterial lipopeptides while TLR4 and TLR5 are receptors for Gram-negative bacterial cell wall components such as LPS or flagellin, respectively. Intracellular TLRs detect viral-derived or synthetic dsRNA and viral-related single-stranded RNA (ssRNA) and unmethylated cytosine-phosphate-guanine DNA (CpG-DNA) of bacteria and viruses (Fig. 1), but ligands of TLR10, TLR12 and TLR13 are still unidentified (Beutler et al., 2006; Beutler, 2009; Seki and Brenner, 2008; Szabo et al., 2006; Takeda and Akira, 2005).

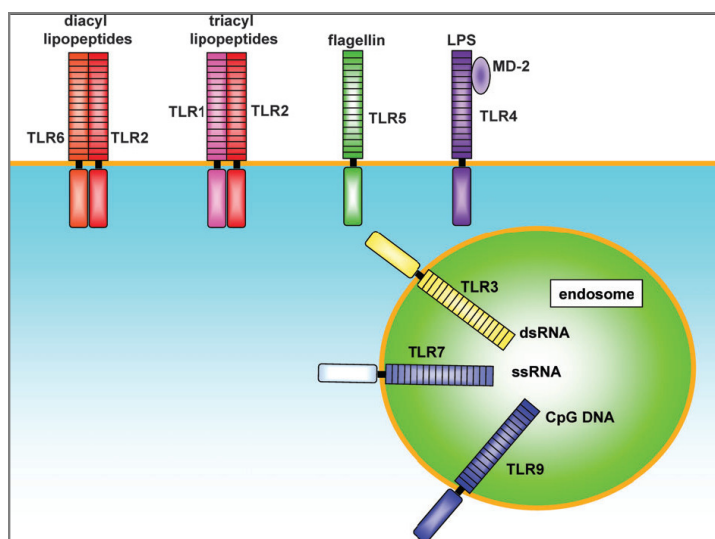


Figure 1: TLRs and their ligands

TLR2 is essential in the recognition of microbial lipopeptides. TLR1 and TLR6 cooperate with TLR2 to discriminate subtle differences between triacyl and diacyl lipopeptides, respectively. TLR4 is the receptor for LPS. TLR9 is essential in CpG-DNA recognition. TLR3 is implicated in the recognition of viral dsRNA, whereas TLR7 and TLR8 are involved in viral-derived ssRNA recognition (Alexopoulou et al., 2001; Heil et al., 2004; Takeda and Akira, 2005). TLR5 recognizes flagellin (From Takeda and Akira, 2005).

1.1.3. TLR expression in different cell types

The epithelial layer provides the first line of defence against invading pathogens. As a consequence, many TLRs are expressed on epithelial cells of the intestinal, respiratory and urogenital tracts. Microbial detecting by TLRs expressed by these cells leads to production of cytokines, chemokines and antimicrobial peptides (Iwasaki and Medzhitov, 2004). Many types of PRRs are expressed by dendritic cells (DCs), including C-type lectins, mannose receptors and TLRs. Studies of DC subsets revealed that TLRs have distinct expression patterns and there are also distinct differences in expression and responsiveness of certain TLRs in freshly isolated DCs versus *in vitro*-derived DCs. Plasmacytoid DCs (pDCs) express TLR7 and TLR9, whereas myeloid DCs (mDCs) express TLR1-3, TLR5, TLR6 and TLR8. Human blood monocytes express TLR1, 2, 4 and 5. Only monocytes and *in vitro*-differentiated DCs express TLR4 and respond thus to LPS stimulation (Iwasaki and Medzhitov, 2004).

TLRs are also expressed by liver cells. Hepatic pDCs also produce tumor necrosis factor (TNF) α , interleukin (IL) -6 and IL-12 by TLR7 and TLR9 stimulation. In general, hepatic DCs are hyperresponsive to TLR ligands. Kupffer cells are the first cells to encounter gut-derived toxins including LPS, thus TLR4 signalling drives Kupffer cells to produce TNF- α , IL-6, -12, -18 and anti-inflammatory cytokine IL-10. IL-12 and IL-18 activate hepatic natural killer (NK) cells to increase the synthesis and release of antimicrobial IFN- γ . Kupffer cells also express TLR2, TLR3 and TLR9. Hepatocytes express a very low level of TLR2-5 and their responses are fairly weak *in vivo*. Hepatic stellate cells (HSCs) express TLR4 and TLR9, while TLR2-5 are expressed by biliary epithelial cells (Seki and Brenner, 2008).

1.1.4. TLR signalling pathways

All TLR members, except TLR3, are associated with the common adaptor molecule MyD88 to trigger inflammatory response such as TNF- α and IL-12. TLR3 and TLR4 utilize the Toll/interleukin-1 receptor (TIR) domain containing the adapter-inducing IFN- β (TRIF) molecule to induce type I IFN. TLR signalling pathways are MyD88-dependent or MyD88-independent/TRIF-dependent (Fig. 2). The MyD88-dependent pathway leads to activation of the transcription factor NF- κ B and induces expression of inflammatory cytokines like TNF- α , IL-6 and IL-12. TLR3 and TLR4 utilize the MyD88-independent component to induce IFN- β . Here, TRIF is used to activate an inhibitor of NF- κ B, which leads to activation of IRF-3 followed by IFN- β production (Armant and Fenton, 2002; Seki and Brenner, 2008; Szabo et al., 2006; Takeda and Akira, 2005).

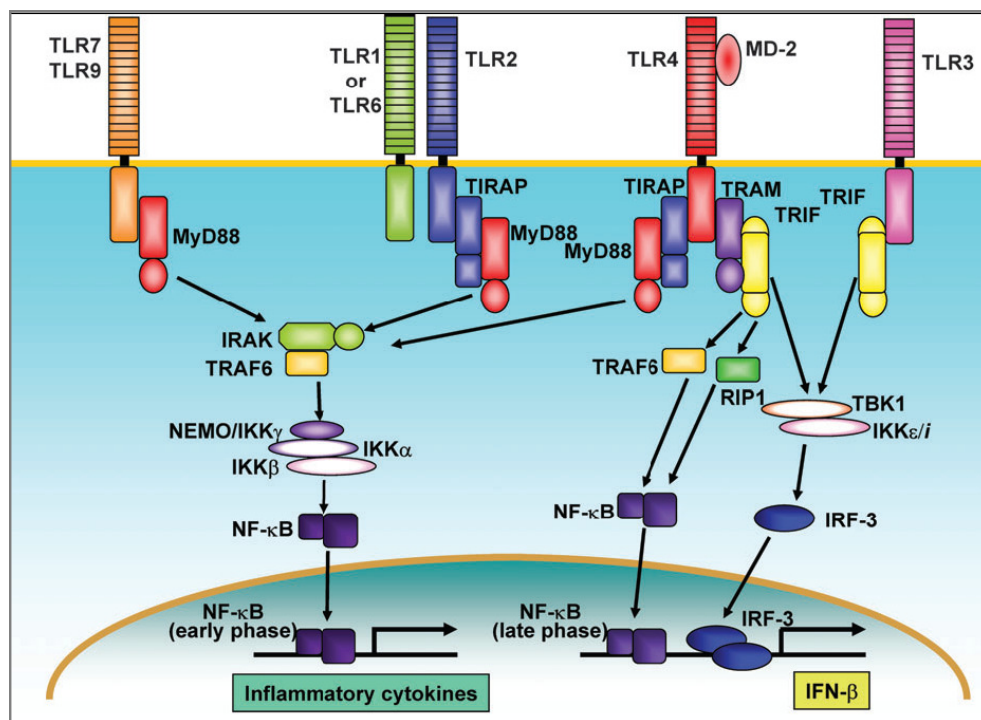


Figure 2: TLR signalling pathways

TLR signalling pathways originate from the cytoplasmic TIR domain. A TIR domain-containing adaptor, MyD88, associates with the cytoplasmic TIR domain of TLRs, and recruits IL-1 receptor-associated kinase (IRAK) to the receptor upon ligand binding. IRAK then activates tumor necrosis receptor-associated factor (TRAF)-6, leading to the activation of the inhibitor of NF κ B (I κ B) kinase (IKK) complex which consists of IKK α , IKK β and NF κ B essential modulator (NEMO)/IKK γ . The IKK complex phosphorylates I κ B and results in nuclear translocation of NF- κ B which induces expression of inflammatory cytokines. A second TIR domain containing adaptor protein (TIRAP), is involved in the MyD88-dependent signalling pathway via TLR2 and TLR4. In TLR3- and TLR4-mediated signalling pathways, activation of IRF-3 and induction of IFN- β are observed in a MyD88-independent manner. A third TIR domain-containing adaptor, TRIF, is essential for the MyD88-independent pathway. Non-typical IKKs, IKK ϵ /IKK δ and the threonine protein kinase 1 (TBK1) mediate the activation of IRF-3 downstream of TRIF. A fourth TIR domain-containing adaptor molecule (TRAM) is specific to the TLR4-mediated MyD88-independent/TRIF-dependent pathway (From Takeda and Akira, 2005).

1.1.5. TLR4-MD2-CD14 complex and LPS

LPS is recognized by a receptor complex of TLR4, myeloid differentiation protein 2 (MD2) and the cluster of differentiation protein 14 (CD14), which are present on many mammalian cell types including macrophages and DCs (Miller et al., 2005; Seki and Brenner, 2008). Recognition of lipid A (endotoxin), a LPS component (Fig. 3a), requires an accessory protein, the LPS-binding protein (LBP), which delivers LPS to CD14, a high-affinity, glycosyl phosphatidylinositol (GPI)-anchored membrane protein that can also circulate in a soluble form. CD14 concentrates LPS for binding to the TLR4-MD2 complex (Fig. 3b). Upon binding of LPS, the intracellular domain of TLR4 recruits TIRAP and MyD88 for MyD88-dependent signalling, resulting in activation of NF- κ B, while TRAM bridges TRIF for the MyD88-independent pathway to activate IRF-3 which induces type I IFN expression (Fig. 2) (Beutler, 2009; Godowski, 2005; Palsson-McDermott and O'Neill, 2004; Seki and Brenner, 2008).

Although lipid A is an essential component of all Gram-negative bacterial membranes, it is highly diverse. Number and length of fatty-acid sides of lipid A were observed to be manifold and this variability could have profound implications for disease, particularly in humans, owing to an altered recognition by the TLR4 complex. In fact, structure-function analyses have indicated that number and length of acid side chains are critical for TLR4 signalling in humans (Miller et al., 2005). Humans evolved discriminatory mechanisms for different lipid A structures which could result in different infection disease outcomes. It was shown that the extracellular domain of TLR4 is less conserved than the cytoplasmatic signalling domain and that there is also a hypervariable extracellular domain (Fig. 3b) that can discriminate between different lipid A structures (Miller et al., 2005).

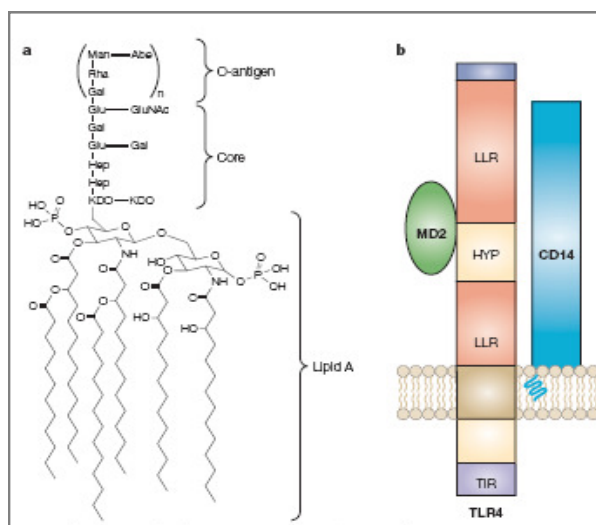


Figure 3: Structure of LPS and TLR4

(a) LPS is composed of lipid A, core oligosaccharide and O-antigen. (b) Components of the TLR4–MD2–CD14 receptor complex. Different TLR4 regions are shown: leucine-rich repeats (LLRs), a hypervariable region (HYP) and the intracellular TIR domain. Membrane bound CD14 with a GPI-anchor (From Miller et al., 2005).

1.1.6. CD14

CD14 is present on the surface of different myeloid cells and at very low numbers on B-lymphocytes, basophils, mammary cells, placental trophoblasts and gingival fibroblasts. Although CD14 is not a transmembrane protein, but is attached to the plasma membrane via the GPI tail, there exists also a soluble CD14 (sCD14) molecule which is found in normal human serum and in culture supernatants of human monocytes and cell lines (Antal-Szalmás, 2000). The sCD14s have an important role in LPS-mediated activation of CD14-negative (but TLR4-positive) cells, like epithelial and smooth muscle cells (Arroyo-Espliguero et al., 2004; Pálsson-McDermott and O'Neill, 2004). Differentiation of monocytes, in tissues, into macrophages is accompanied by a change in membrane CD14 (mCD14) numbers. The synthesis and expression of CD14 in monocytes, macrophages and neutrophils can be altered and regulated by several mediators. In neutrophils, TNF- α and LPS are known to rapidly upregulate CD14 expression, while in monocytes, expression is decreased by the anti-inflammatory cytokines IL-4 and IL-13 (Antal-Szalmás, 2000).

As mentioned before, CD14 is part of the multiprotein complex TLR4-MD2-CD14 and thus acts as a pattern recognition molecule that has a central role in innate immunity. It can interact in a non-specific fashion with several ligands, including LPS from Gram-negative bacteria and components from Gram-positive bacteria and fungi. Both mCD14 and sCD14, together with LBP, are the first line screeners of microbial antigens and present them to the more pathogen-specific signalling receptor TLR4-MD2 (Arroyo-Espliguero et al., 2004; Kitchens and Thompson, 2005).

1.1.7. IRF-1

The IRF family was identified in the late 1980s and consists of nine members in mammals. This family was the first identified transcriptional regulators of type I IFN and IFN-inducible genes and is now recognized as playing a critical part in innate and adaptive immunity (Honda and Taniguchi, 2006; Tamura et al., 2008). Many IRF members (IRF-1, IRF-3, IRF-5 and IRF-8) play a central role in cell growth, cell survival and oncogenesis, as well as in the regulation of gene expression in response to pathogen infection (IRF-1, IRF-3, IRF-4, IRF-5, IRF-7 and IRF-8) (Tamura et al., 2008). In fact, IRF-1 is suggested to provide a link between the innate and adaptive immune system. First of all, IRF-1 induces many genes involved in the initial reaction to pathogen infection such as 2'-, 5'- oligo adenylate synthetase (OAS), protein kinase R (PKR) and the inducible nitric oxide synthase (iNOS). IRF-1 also influences development and function of NK cells and, in addition, IRF-1 overexpression induces IFN- α and IFN- β transcription. These cytokines have stimulatory effects on macrophages and T cells (Kroger et al., 2002).

IRF-1 can also be activated by type I IFN (IFN- α and IFN- β) and type II IFN (IFN- γ). IRF-1 is constitutively localized in the nucleus and the consensus-binding sequence motif has been determined to be 5'- AANNGAAA -3' (the consensus recognition sequence for each IRF) that appears within the promoter of several IRF-inducible genes (Kroger et al., 2002; Tamura et al., 2008).

1.2. RNA polymerase II (RNAPII) transcription cycle

The carboxyl-terminal domain (CTD) of the largest subunit of the RNAPII is essential for transcription, because phosphorylation of the CTD was found to be a key event during messenger RNA (mRNA) metabolism. Phosphorylation of the CTD varies during the transcription cycle and several kinases that are capable of phosphorylation of the CTD have been discovered. The CTD consists of multiple repeats of a seven-amino-acid motif which is in turn made up of tyrosine (Tyr), serine (Ser), proline (Pro) and threonine (Thr). These amino acids form the consensus heptapeptide Tyr¹-Ser²-Pro³-Thr⁴-Ser⁵-Pro⁶-Ser⁷ (Chapman et al., 2007; Palancade and Bensaude, 2003; Prelich, 2002).

The functions of the CTD are quite complex as it is involved in all major steps of mRNA formation, including transcription initiation and elongation, capping, splicing, and 3' end processing (Fig. 4). A variety of kinases have been identified with phosphorylation activity mainly towards the amino acids serine 2 and serine 5 (Chapman et al., 2007; Prelich, 2002). In general, during mRNA metabolism, the unphosphorylated CTD interacts with a mediator complex (Med) (Fig. 4a) to form a preinitiation complex (Fig. 4b). Phosphorylation of the CTD is required to disrupt these interactions at elongation of transcription (Fig. 4d) and to assist the recruitment of pre-mRNA modification enzymes (Fig. 4f). Serine 5 is phosphorylated first in the initiation complex (Fig. 4c) by the cyclin-dependent kinase (CDK) 7, and serine 2 of the CTD phosphorylation is mediated by CDK9 upon entry into elongation (Palancade and Bensaude, 2003). Thus, differential phosphorylation of serine 2 and serine 5 at the 5' and 3' regions of genes appear to coordinate the localization of transcription- and RNA-processing factors to the elongation polymerase complex (Chapman et al., 2007).

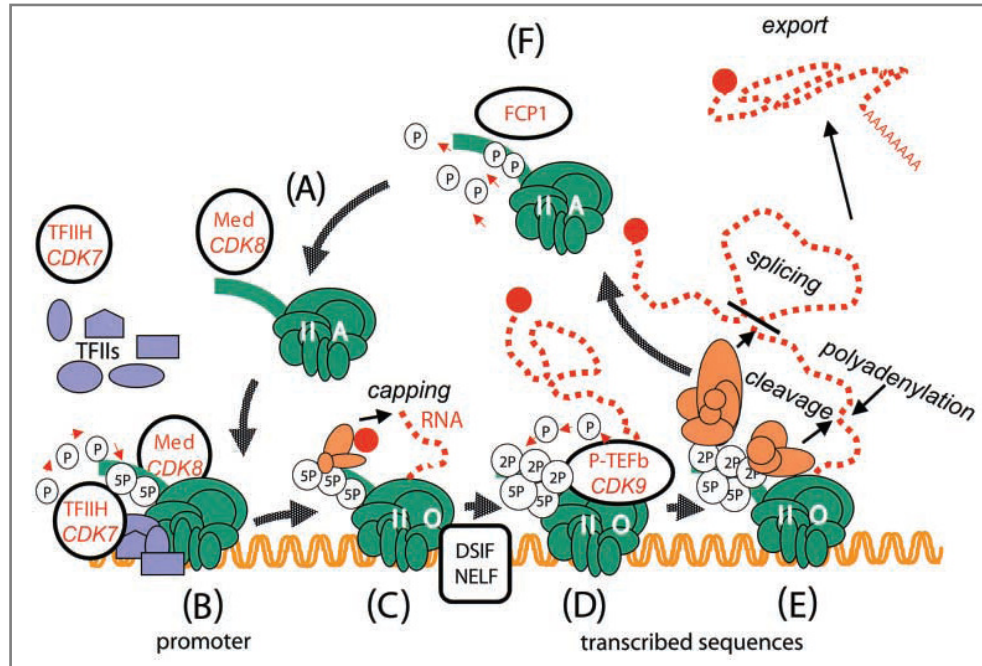


Figure 4: CTD phosphorylation during the transcription cycle

(A) Recycling. The RNAPII core enzyme is not phosphorylated on the CTD. It may assemble with co-activators such as Med to form a holoenzyme. Premature CTD phosphorylation by CDK8 prevents the assembly of RNAPII on the promoter. (B) Preinitiation. The unphosphorylated RNAPII assembles onto the promoter sequences with general transcription factors, thus forming a preinitiation complex of transcription. The CTD is phosphorylated on serines at position 5 (5P) by the CDK7 subunit of general transcription factor II (TFII) H. (C) Initiation. Transcription begins. The phosphorylated CTD recruits the capping enzymes and the nascent transcript is capped at its 5' end. (D) Elongation. Phosphorylation of the CTD on serines at position 2 (2P) by the CDK9 subunit of the positive transcription elongation factor b (P-TEFb) is required to elongate transcription. (E) RNA processing and transcription termination. The phosphorylated CTD recruits the splicing machinery to remove introns and, finally, recruits the cleavage and polyadenylation factors that cleave the transcript and add a polyadenylic tail at its 3' end. This step signals transcription to terminate and RNAPII falls off its DNA template. The resulting mRNA is then exported to the cytoplasm. (F) To be recycled for another transcription round, RNAPII is dephosphorylated by the TFIIH-dependent CTD phosphatase 1 (FCP1) (Modified from Palancade and Bensaude, 2003).

1.3. Single nucleotide polymorphisms (SNPs)

With completion of the Human Genome Project, a large number of subtle variations – polymorphisms - have been found (Kim and Misra, 2007). The most common type of genetic variations is the single nucleotide polymorphism (SNP) with more than 10 million reported in public databases, so far (Kubistova et al., 2009). A SNP is a single variation at a specific location in the genome that is by definition found in more than 1% of the population. Typically, genotypes of these polymorphisms are biallelic, very rarely tri- or tetraallelic, and the average overall frequency of SNPs in the human genome is approximately one per 1,000 bp (Kim and Misra, 2007).

Within a population, a minor allele frequency (MAF) can be assigned for each SNP as a ratio of chromosomes in the population carrying the less common variant to those with the more common variant. There are marked differences between human populations in terms of the distribution of particular SNP variants, thus an SNP allele that is common in one geographical or ethnic group may be rarer in another (Kubistova et al., 2009).

It was found that SNPs do occur less frequently in coding regions of the genome than in non-coding regions. SNPs, when present in regulatory sites of a gene, can affect rates of transcription causing changes in the production of encoded proteins or may have consequences for transcript splicing/stability, transcription factor binding or the sequence of non-coding RNA. On the other hand, SNPs located within coding regions can cause alterations in protein structure or function and are thus called non-synonymous. SNPs within a coding region will not necessarily change the amino acid sequence of the protein, due to degeneration of the genetic code, so polymorphisms in which both forms lead to the same polypeptide sequence are termed synonymous (Kim and Misra, 2007; Kubistova et al., 2009).

SNPs that have an effect on gene regulation, e.g. a functional effect on gene expression, are said to be regulatory SNPs (rSNPs). The rSNPs are potentially susceptibility SNPs for phenotypes, including diseases (Buckland, 2006). Regulation at the level of transcription initiation and RNA processing defines downstream biological effects. Such regulation occurs in *cis*, directly affecting the regulated gene, but it can act in *trans* by altering activity of downstream genes. Significant interindividual differences in gene expression patterns are common and may result from both environmental factors and *cis*- or *trans*-mediated genetic effects (Johnson et al., 2005).

SNPs in immune response genes (e.g. TLRs, cytokines, or immunoglobulins) may have severe consequences, as these immune molecules are implicated in the pathogenesis of a large number of human diseases. Thus variations in the DNA sequence of these genes can affect whether or how humans develop diseases and respond to pathogens, chemicals or drugs (Kubistova et al., 2009).

The most commonly studied polymorphisms, nonsynonymous changes that alter amino acid coding, appear in many cases insufficient to account for interindividual differences in disease aetiology and response to therapies. Further, it is estimated that functional polymorphisms that are *cis*-regulatory in the human genome outnumber those that alter protein sequence, and that the bulk of regulatory polymorphisms remain to be discovered (Johnson et al., 2005).

1.3.1. Methods for investigation of rSNPs

Investigation of SNPs that might cause changes in cellular biological processes that induce disease states is performed in disease genetics studies. A typical approach are case-control association studies, where genotyping in a patient group and healthy control population is performed to investigate relationships between specific genotypes (SNPs) and phenotypic characteristics (Kim and Misra, 2007; Kubistova et al., 2009).

The reporter gene assay is a widely used and very sensitive method for the study of promoter strength. Briefly, the promoter is cloned directly upstream of the reporter gene in a promoterless vector and introduced into cultured cells. The reporter gene is expressed in an extra chromosomal state and the protein activity can be quantified. Thus, the effect of isolated SNPs can be assayed. But, because the promoter is not in its natural chromatin environment and because of the different behaviour of cells in culture, the results are sometimes difficult to correlate to *in vivo* observations (Pampin and Rodriguez-Rey, 2007).

The electrophoretic mobility shift assay (EMSA) is a major method for detecting DNA/protein interactions. A labelled double-stranded oligonucleotide is mixed with a nuclear extract prepared from cells that express transcription factors. The formation of DNA/protein complexes result in a retardation of mobility during electrophoresis and, thus, a separation from the free probe. The effect of an rSNP on transcription factor binding results in changes of electrophoretic mobility (Pampin and Rodriguez-Rey, 2007).

Quantification of gene expression is another approach of functional studies to investigate rSNPs. These analyses rely on comparisons of the relative abundance of the mRNA between genotypic groups (Buckland, 2006).

To circumvent intrinsic errors in interindividual comparison and to analyse allele-specific gene expression within its natural genomic environment, two techniques have been developed more recently: allele-specific transcript quantification (ASTQ) and haplotype-specific immunoprecipitation (HaploChIP).

For the first technique, a polymorphism in the transcribed region of a gene is used as a marker to distinguish between the allelic mRNA products of genes harbouring a putative regulatory polymorphism within their regulatory regions in one double heterozygous individual. The relative abundance of allelic mRNAs from a heterozygous individual is then quantified in allele-specific polymerase chain reactions (PCR). Both gene copies originate from the same tissue sample and identical environmental influences, so each allele should be equally expressed regardless of the absolute level of gene expression. The ratio of the abundance of each allele is therefore expected to be ~1. In samples that are heterozygous for a regulatory variant, mRNA originating from one chromosome will be expressed at a higher level than that from its sister chromosome and this is detected by changes in the ratio of the amount of each mRNA allele (Buckland, 2006).

The second approach to measure allele-specific transcription *in vivo*, HaploChIP, investigates RNAPII binding to genomic DNA (gDNA) using an SNP in the promoter region as a marker. Here an antibody specific to serine 5 phosphorylated (ser5-P) RNAPII is added to fragments of cross-linked DNA/protein complexes and the immunoprecipitated chromatin is measured for the relative amount of each precipitated allele. This method has the advantage of allowing the functional association of DNA variants in the promoter with changes in gene expression (Buckland, 2006; Knight et al., 2003; Pampin and Rodriguez-Rey, 2007).

1.3.2. Association of the polymorphisms CD14 rs2569190 and IRF-1 rs2549009 with various disease conditions

The CD14 rs2569190 (C/T) polymorphism was first identified by Baldini and colleagues. The homozygous TT genotype of this SNP was described as being associated with higher levels of sCD14 and decreased IgE levels in children who were skin test-positive for local aeroallergens (Baldini et al., 1999). Later, this SNP was found to influence CD14 gene expression via a differential effect on binding of transcription factors Sp1, Sp2 and Sp3 to gDNA in different cell types (LeVan et al., 2001). Also, a gene-environment interaction between T allele carriers and *Helicobacter pylori* antibodies on serum total IgE was observed in Finish children (Virta et al., 2008). Further, the homozygous TT genotype was suggested to confer an increased risk of alcoholic liver disease progression, especially liver cirrhosis in a Finnish population (Jarvelainen et al., 2001), and it was found to be associated with mild cryptogenic chronic liver disease (Von Hahn et al., 2008). The same genotype was later found to be associated with significantly higher sCD14 levels in patients with hepatitis C virus (HCV) infection (Meiler et al., 2005). Another study reported a possible relationship between CD14 rs2569190 TT genotype and the formation of portal lymphoid aggregates in chronic HCV infection (Askar et al., 2009).

The SNP rs2549009 (G/A) is located within the promoter region of the transcription regulator gene IRF-1. The homozygous genotype AA was associated with the outcome of HCV genotype 3 infection (Wietzke-Braun et al., 2006). This genotype was also positively associated with Th1-type response in patients with chronic HCV infection, suggesting that this SNP variant might be beneficial for viral elimination in chronic HCV and IFN therapy (Saito et al., 2002).

Additionally, it was shown that DNA constructs, containing the genetic polymorphism rs2549009, modified promoter activity in luciferase promoter assays (Saito et al., 2001; Schedel et al., 2008).

1.4. Aims of this work

The aim of this study is to determine whether the polymorphisms rs2569190, located within the CD14 gene, and rs2549009, located within the promoter region of the IRF-1 gene, confer any regulatory effect on transcriptional activity in their natural genomic context in human peripheral blood mononuclear cells (PBMC). For this purpose the technique HaploChIP, which assesses the relative binding frequency of RNAPII to allelic gene promoter regions at different stages of the transcription cycle by using RNAPII-specific antibodies, was established and applied.

2. Materials & Methods

2.1. Blood donors

For cross-linking of proteins to the DNA in PBMC (during the HaploChIP method see 2.3.) blood samples from a cohort of 13 healthy Caucasians (6 female, 7 male; mean age 30.0 years, range 23-48) were taken. PBMC were prepared by Ficoll density centrifugation and isolated, and cross-linked cells were stored at -80°C for further analyses. CD14 rs2569190 and IRF rs2549009 heterozygotes (n=9; n=6) were identified (see 2.2.) and asked to give further blood samples of 30 ml venous blood for immunoprecipitation during HaploChIP. Experiments were performed at least two times each for independent replication. Later, these 13 donors were included in a larger cohort consisting of 42 healthy Caucasians which was used for further functional and association studies by our research team (i.e. IRF-1 gene expression with regard to IRF-1 rs2549009 genotypes).

2.2. Genotyping

From whole blood samples, gDNA was purified using the QIAamp DNA Blood Mini Kit following the blood and body fluid spin protocol (Qiagen, Hilden, Germany). Genotyping was performed using 10-40 ng gDNA. Allelic discrimination of variant positions within the CD14, IRF-1 and the imprinted small nuclear ribonucleoprotein polypeptide N (SNRPN) gene was achieved by different SNP-specific assays (Tab. 1).

Tab. 1: Genotyping of variant positions using different SNP-specific assays

gene and SNP #	assay type	reference
CD14 rs2549190	5'-nuclease assay (custom designed)	(Maouzi, 2007)
CD14 rs5744455	allele-specific tetra-primer amplification refractory mutation system- (ARMS-) PCR	(Bregadze, 2010; Mertens et al., 2009)
CD14 rs2563298		
IRF-1 rs2549009	5'-nuclease assay (custom designed)	(Maouzi, 2007)
IRF-1 rs2549007	allele-specific primer pairs	this thesis
IRF-1 rs2706384		
SNRPN rs705	5'-nuclease assay (ABI genotyping assay C_2066555_10)	(Mertens et al., 2009)

2.2.1. Discrimination of variant positions using 5'-nuclease assays

Genotyping of variant positions by allelic discrimination in 5'-nuclease assays was carried out in the sequence detection system ABI StepOnePlus or ABI prism 7000 (Applied Biosystems, Darmstadt, Germany) according to supplier's instructions. For genotyping of CD14 rs2549190 and IRF-1 rs2549009, custom designed 5'-nuclease assays were available from an earlier thesis. Discrimination of gDNA and transcript variants at SNP rs705 within the imprinted gene SNRPN was carried out using a commercially available genotyping assay (Applied Biosystems).

2.2.2. Discrimination of variant positions using ARMS-PCR

Amplification of ARMS-PCR via tetra-primers was performed using conventional PCR followed by analysis of PCR products in agarose gels. This method uses two outer primers and two inner primers, which are specific for the variant position of interest. One outer primer and one inner primer each are specific for the variant alleles of SNP, while both outer primers together generate a non-SNP- but gene-specific PCR product. Thus, it is possible to generate allele-specific as well as non-allele-specific amplicons in one reaction. Tetra-primers were designed using a software publicly provided by Ye and colleagues (Ye et al., 2001).

Amplification CD14 rs5744455 was performed using 10-20 ng gDNA in a total of 25 µl GoTaq Master Mix (GoTaq Green-/Colorless- Master Mix, Promega GmbH, Mannheim, Germany) with 0.18 µM primers each (rs574 tetra T f 5'-AAG GAA GGG GGA ATT TTT CTT TAG CCT-3', rs574 tetra C r 5'-GGT AGA ATT AGG TTC AAG AAA AGG AAG GTG-3', rs574 tetra for 5'-CTG AGC AAC AGA GCA AGA CTC TAT CTC A-3' and rs574 tetra rev 5'-CAT TTC TTT GAC TTC TTC CTT GTC TTG G-3'). Cycling conditions were 3 min initial denaturation at 94° C, 40 cycles of 1 min denaturation at 94° C and 1 min annealing at 62° C and 1 min elongation at 72° C.

Genotyping of CD14 rs22563298 was performed similarly with 0.18 µM primers each (rs2563 tetra G f 5'-CCC ACC TTT ATT AAA ATC TTA AAG AAC AGG-3', rs2563 tetra T r 5'-TAA AGG TCT GTT AAA TGA ATG ACA CGA AA-3', rs2563 tetra for 5'-TTT GCC TAA GAT CCA AGA CAG AAT AAT G-3' and rs2563 tetra rev 5'-TTT CTT AGG GAG TTA GGA TGA AGA AAG C-3') using identical cycling conditions except annealing at 60° C.

For CD14 rs5744455 products of 277 bp for the outer primers, 142 bp for the C allele and 192 bp in size for the T allele were generated (Fig. 5a). PCR products in size of 247 bp for outer primers, 143 bp for the G allele and 163 for the T allele were found for genotyping of CD14 rs2563298 (Fig. 5b). Thus, product patterns of 277/142, 277/142/192 and 277/192 bp correspond to CC, CT and TT genotypes of CD14 rs5744455. ARMS-PCR patterns of 247/143, 247/143/163 and 247/163 bp correspond to CD14 rs2563298 GG, GT and TT genotypes (Mertens et al., 2009).

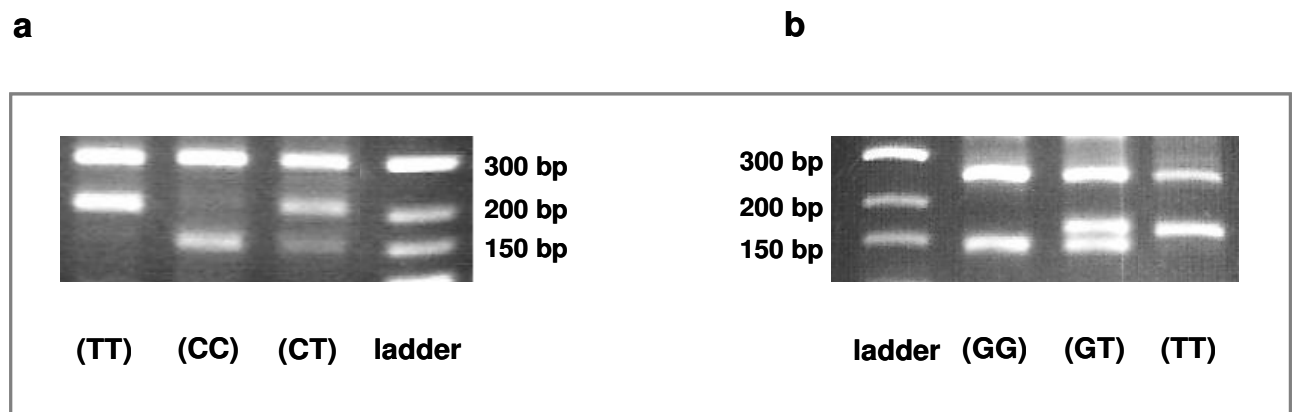


Figure 5: Discrimination of CD14 rs5744455 and CD14 rs2563298 using ARMS-PCR – agarose gel

For analysis of PCR products for the variant position CD14 rs5744455 (a) and CD14 rs2563298 (b) agarose gels of 2% and 4 % were used, respectively. Amplicon size patterns indicate genotypes as outlined in the text.

2.2.3. Discrimination of variant positions using allele-specific primers

Discrimination of the polymorphisms rs2549007 and rs2706384 within the IRF-1 promoter region was achieved on gDNA from the larger cohort of healthy donors in real-time PCR using allele-specific primers. For genotyping of the variant positions, primers were used that are able to distinguish between different alleles of the genetic polymorphisms. Therefore, allele-specific primers were designed modelled on tetra-primers, but applicable for real-time PCR (Ye et al., 2001). Unlike ARMS-PCR, which also generates non-SNP-gene-specific PCR products for outer primers in this real-time PCR approach, only allele-specific products originate depending on the genotype for the respective donor (Fig. 6).

IRF-1 rs2549007 genotyping was performed using 2 ng gDNA in a total volume of 10 µl using 0.30 µM of primers each (IRF-1 rs254 for 5'-AGG CTT TCT GCC TTC TTC ACT T-3' and IRF-1 rs254 G r 5'-GTA TAT CTC CCG AAC GCA TGC-3', or IRF-1 rs254 A r 5'-GGT ATA TCT CCC GAA CGC ATG T-3') in standard SYBR green reactions (Platinum SYBR GREEN qPCR SuperMix-UDG, Invitrogen GmbH, Darmstadt, Germany). Cycling conditions were 2 min initial denaturation for activation of the antibody blocked *Taq* DNA polymerase at 50°C, followed by 2 min denaturation at 95°C and 45 cycles of 15 s at 95°C denaturation and 30s at 60°C annealing and extension.

Genotyping of IRF-1 rs2706384 was also performed in standard SYBR green (Invitrogen) amplification reactions of 4 ng gDNA using 75 nM of primer pairs each (IRF-1 rs270 rev 5'-CTG TCC TCT CAC TCC GCC TTG TCC-3' and IRF-1 rs270 C f 5'-CAA GTG CCC GGG CGA CCC-3', or IRF-1 rs270 A f 5'-CAA GTG CCC GGG CGA CCA-3'). Cycling conditions were equal, but an annealing step at 67°C for 30 s was necessary.

Specificity of PCR products was validated by melting curve analyses. In addition to the allele-specific PCR products, non-specific amplicons were also generated to a rather lesser amount (> 19-fold difference), most probably due to unspecific binding of primers. Thus, the different genotypes GG, AG or AA for IRF-1 rs2549007 and CC, AC or AA for IRF-1 rs2706384 were assigned due to the very predominant amplification of the one or the other allele or to a rather equal amplification of both alleles, depending on the genotype for the respective donor (Fig. 6).

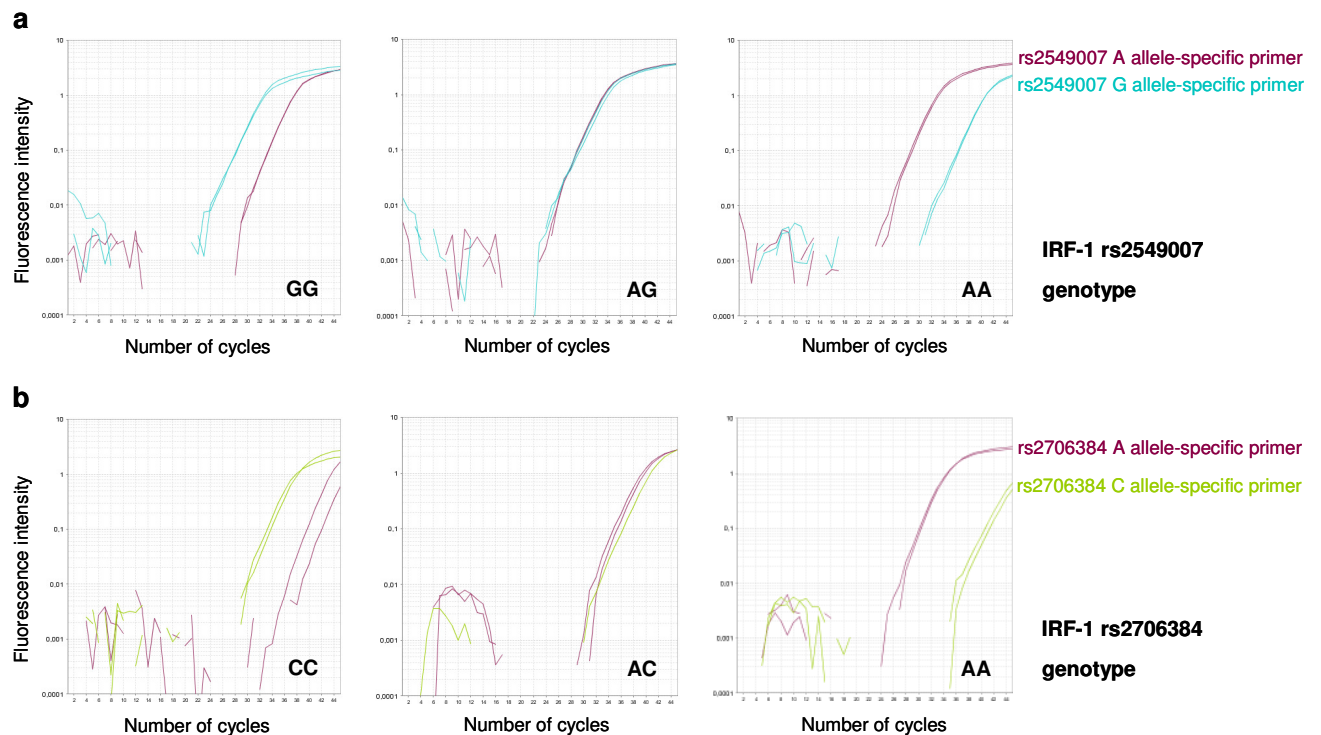


Figure 6: Genotyping of IRF-1 rs2549007 and IRF-1 rs2706384 using allele-specific primers in real time PCR

Discrimination of variant positions within the promoter region of IRF-1 was achieved in real-time PCRs using allele-specific primer for (a) IRF-1 rs2549007 and (b) IRF-1 rs2706384. In general, allele-specific primers generated PCR products with fluorescence signals of different intensities representing specific and unspecific binding of primers, depending on the respective genotype. Genotypes GG, AG or AA for IRF-1 rs2549007 (a) and CC, AC or AA for IRF-1 rs2706384 (b) were determined by predominant or equal amplification of the variant alleles, respectively.

2.3. HaploChIP

HaploChIP analysis comprised (a) cross-linking of proteins to the DNA and enzymatic shearing of chromatin, (b) an immunoprecipitation of chromatin fragments by antibodies with different specificities, and (c) a restriction endonuclease digestion of amplified chromatin fraction.

2.3.1. Cross-linking and enzymatic shearing of chromatin

Chromatin immunoprecipitation was conducted using the ChIP-IT Express kit (Active Motif, Rixensart, Belgium) according to the supplier's protocol. In brief, $1-10 \times 10^6$ freshly isolated PBMC were cross-linked using 1% formaldehyde for 5-6 min (Fig. 7a). After isolation of nuclei (Fig. 7b) enzymatic digestion reactions were performed using a shearing cocktail for about 8-10 min at 37°C (Fig. 7c). Sheared chromatin was then stored in aliquots at -80°C until further use for immunoprecipitation or as control 'input' DNA.

2.3.2. Immunoprecipitation

Sheared chromatin corresponding to 1×10^6 cells was subjected to immunoprecipitation by 4 µg of an (i) IgG-antibody directed against the transcriptionally inactive, ser5-P RNAPII (39097 clone 4H8, Active Motif) or by a control mock antibody (53010 negative control mouse IgG 101226, Active Motif) (Fig. 7d). Antibody-bound protein/DNA complexes were captured through use of magnetic protein G-coated beads.

(ii) Immunoprecipitation was also performed by IgM-antibodies directed against ser5-P (MMS-134R clone H14, Covance, California, USA) or the transcriptionally active, serine 2 phosphorylated (ser2-P) RNAPII (MMS-129R clone H5, Covance) or by an IgM-isotype (ITC0692 clone A16G8, Linaris GmbH, Wertheim-Bettingen, Germany) using 5 ng of an bridging anti-mouse-IgM antibody (315-005-020, Jackson ImmunoResearch, West Grove, Pennsylvania, USA) (Fig. 7d). The additional bridging step was necessary to allow immunoprecipitations by IgM antibodies and magnetic beads, which are specific for IgG antibodies because of its protein G coat. Bridging was performed by pre-treatment of 25 µl protein G magnetic beads with 5 µg of the anti-IgM antibody for 1 h at 4° C according to supplier's instructions.

Captured chromatin was then subjected to reverse-cross-linking and protease K digestion (Fig. 7e).

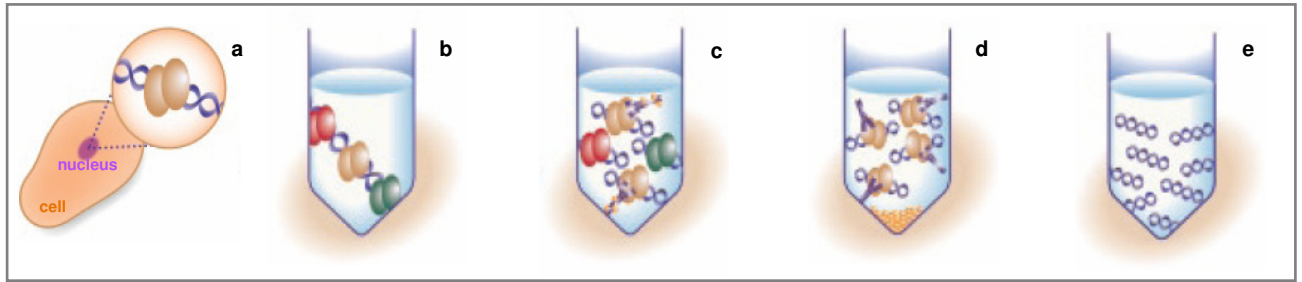


Figure 7: Schematic overview of the HaploChIP technique

The principle procedure of the HaploChIP technique taken from Version G (modified) of ChIP-IT™ Instruction Manual (Active Motif) is shown. After cross-linking of proteins to the DNA (a), isolated chromatin (b) was subjected to enzymatic shearing (c). Sheared chromatin fragments were then immunoprecipitated by different antibodies (d) and subjected to reverse-cross-linking and digestion with protease K (e).

2.3.3. Restriction fragment length polymorphism (RFLP) analysis

RFLP analysis included amplification of regions around CD14 rs2569190 or IRF-1 rs2549009, followed by SNP-dependent digestion. Non- and precipitated chromatin fragments (Fig. 7e) were used as templates for low cycle number or standard cycle number PCR reactions spanning CD14 rs2569190 or IRF-1 rs2549009. PCR products were then subjected to digest reactions using respective SNP-specific endonucleases.

2.3.3.1. Amplification of purified chromatin fragments

Amplification spanning a region of 201bp around CD14 rs2569190 was performed using 8 µl captured and 0.8 µl non-captured material in a total of 100 µl GoTaq Master Mix (Promega GmbH) with 0.30 µM primers each (rs2569190 for 5'-GAC ACA GAA CCC TAG ATG CCC T-3' and rs2569190 rev 5'-GTG AAC TCT TCG GCT GCC TC-3'). Cycling conditions for (i) low cycle number PCR were 2 min initial denaturation at 94° C, 24-28 cycles of 30 s denaturation at 94° C and 1 min annealing and elongation at 60° C, followed by a final elongation step for 7 min at 60° C. (ii) Standard cycle number PCRs (40-45 cycles) were run under the same conditions.

Amplification spanning IRF-1 rs2549009 (142bp) was performed using 4 µl antibody-captured and 0.4 µl non-immunoprecipitated material in a total of 50 µl GoTaq Master Mix (Promega GmbH) with 0.30 µM of primers each (IRF rs2549 for 5'-CAG GAG GGT GAA AAG ATG GCC-3' and IRF rs2549 rev 5'-CGA GCG CTC CCA ATC CAC-3'). Cycling conditions were 3 min initial denaturation at 94° C, 45 cycles of denaturation at 94° C for 30 s, annealing for 30 s at 60° C and 72° C elongation for 30 s, followed by a final elongation step at 72° C for 10 min.

Size of PCR products was determined in 1% agarose gels. Because of unspecific products, all PCR reactions were purified and concentrated by gel extraction using the PrepEase gel extraction kit (USB, Cleveland, OH, USA) according to supplier's instructions.

2.3.3.2. Restriction endonuclease digestion of amplicons

SNP-specific endonucleases were used for allelic digestion of PCR products. Therefore, detection of highly specific enzymes was achieved using the software NEBcutter V2.0 (Vincze et al., 2003). For allelic digestion of CD14 PCR products, a region of 41 nt around CD14 rs2569190, 20 nt upstream and downstream, was analyzed for C allele- (Fig 8a) and T allele- (Fig 8b) specific endonucleases. The enzyme *Hae*III was found to recognize a DNA sequence containing CD14 rs2569190 C, while *Hpy*CH4III (*Taa*I) was found to recognize a DNA sequence containing the T allele.

Identification of suitable endonucleases for IRF-1 rs2549009 was performed similarly. Amongst others, the enzymes *Nci*II (*Bcn*I) and *Bst*NI (*Mva*I) were indicated by the software (Fig 9) to be specific for a DNA sequence containing the IRF-1 rs2549009 wild-type allele G or variant allele A (for detailed recognition sequences see 5.3.).



Figure 8: Detection of CD14 rs2569190 C and T allele-specific endonucleases using the NEBcutter V2.0

A region of 20 nt in both directions, upstream and downstream, of CD14 rs2549190 was analyzed for allele-specific endonucleases using the software NEBcutter V2.0. All commercially available endonucleases specific for the shown sequence are indicated. DNA sequences containing (a) the wild-type allele C and (b) the variant T, located at position 21 in the cut-out, are recognized by, amongst others, *Hae*III and *Hpy*CH4III (*Taa*I).

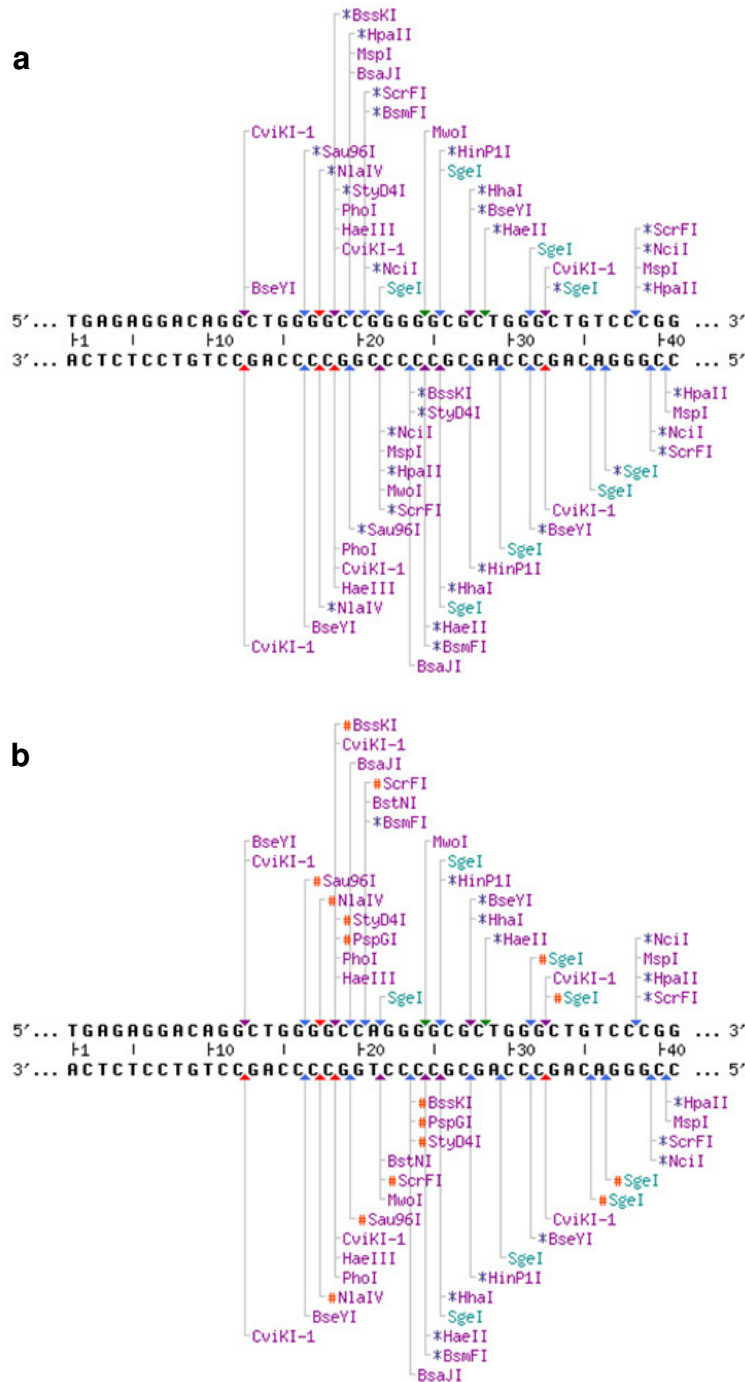


Figure 9: Detection of IRF-1 rs2549009 G and A allele-specific endonucleases using the NEBcutter V2.0

A DNA sequence region of 20 nt upstream and downstream of IRF-1 rs2549009 was analyzed for allele-specific endonucleases using the software NEBcutter V2.0. All commercially available endonucleases specific for the shown sequence are given. DNA sequences containing (a) the wild-type allele G or (b) the variant allele A, located at position 21 in the cut-out, are recognized by *NciI* (*BcnI*) or *BstNI* (*MvaI*).

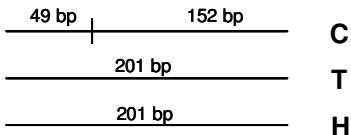
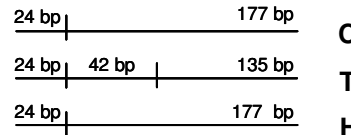
Digestion of CD14 rs2569190 was carried out in a total of 10 µl digestion reactions using 8 µl purified PCR product with 5 U endonucleases, *HaeIII* or *TaaI* (Fermentas GmbH, St. Leon-Rot, Germany) in respective buffers at 37° C or 65° C for at least for 4 h or overnight.

Digestion of IRF-1 rs2549009 was performed similarly, using 5 U endonucleases, *BcnI* or *MvaI* (Fermentas GmbH) in a total of 10 µl and respective buffers at 37° C for 2-3 h.

All digestion reactions were stopped and purified again using the PrepEase gel extraction kit (USB) following the PCR purification protocol. For digestion fragments see Fig. 10.

Because of an excess of PCR products during the saturation stage of the amplification reaction, allelic amplicons might bind to each other independently of the single variant position during annealing, thus forming hybrids composed of one strand containing one allele and a complementary strand carrying complementary nucleotides of the other allele. Theoretically, up to 50% of amplicons might be hybrids. These hybrids could not be recognized and digested by the allele-specific endonucleases (Fig. 10).

a

<i>HaeIII</i>			<i>TaaI</i>		
CC	CT	TT	CC	CT	TT
152 bp 49 bp	201 bp 152 bp 49 bp	201 bp	177 bp 24 bp	177 bp 135 bp 42 bp 2x 24 bp	135 bp 42 bp 24 bp
					
		C			C
		T			T
		H			H

b

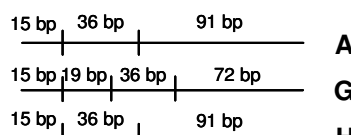
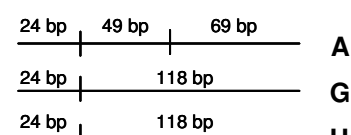
<i>BcnI</i>			<i>MvaI</i>		
GG	GA	AA	GG	GA	AA
72 bp 36 bp 19 bp 15 bp	91 bp 72 bp 2x 36 bp 19 bp 2x 15 bp	91 bp 36 bp 15 bp	118 bp 24 bp	118 bp 69 bp 49 bp 2x 24 bp	69 bp 49 bp 24 bp
					
		A			A
		G			G
		H			H

Figure 10: Digestion fragments

(a) Digestion of CD14 rs2569190 PCR with *HaeIII* or *TaaI* resulted in specific digestion fragments for each reaction. (b) Digestion of IRF-1 rs2549009 PCR with *BcnI* or *MvaI* also resulted in different fragments of specific size for each variant allele. During PCR, hybrid (H) products were also found to be generated in addition to completely homologous double-stranded amplicons. In addition to the respective SNP-specific recognition sequence, SNP-independent DNA sequences within the amplicons were also digested by the restriction enzymes. The relative amount of digested and undigested fragments was then analyzed in a bioanalyzer.

2.3.3.3. Quantification of DNA fragments with a bioanalyzer

All digested fragments were analyzed in the Agilent 2100 bioanalyzer using the DNA 1000 LabChip kit (Agilent Technologies, Böblingen, Germany) according to the supplier's instructions. The bioanalyzer uses a chip-based technology to separate and quantify nucleic acids. DNA fragments are separated according to their length in a gel within the chip and because of different fluorescence intensities, the software is able to calculate respective molarities for each fragment.

1 µl of each cleaned digestion reaction, CD14 rs2569190/*Hae*III or *Taa*I (Fig. 11a) or IRF-1 rs2549190/*Bcn*I or *Mva*I (Fig. 11b), was analyzed for the amount of digested and undigested PCR fragments. The relative amount of the respective variant allele fragments was calculated by given molarities for each fragment and PCR hybrid.

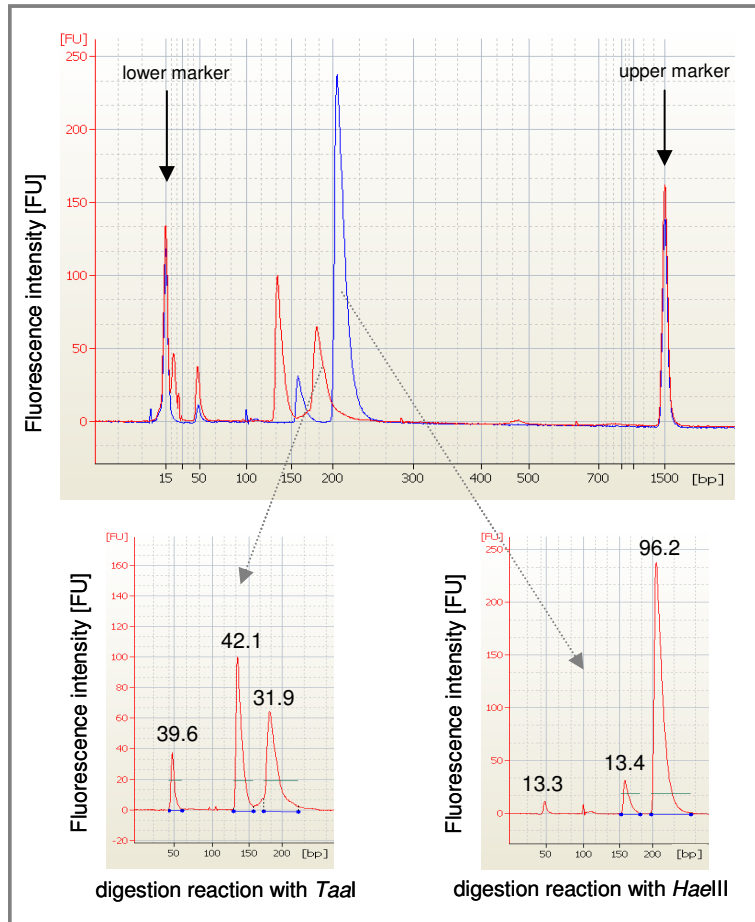


Figure 11a: CD14 rs2569190 RFLP analysis of output material from individual #3 using the Agilent 2100 bioanalyzer

Output material (individual #3), was subjected to amplification spanning a region around CD14 rs2569190 SNP and yielding a product of 201bp in size. Amplicons were subjected to two different restriction enzyme digestions with *TaaI* and *HaeIII*. Digestion with *TaaI* (indicated in red) resulted in fragments of 135bp, 42bp and 24bp in size representing T allele DNA, while fragments of 177bp and 24bp represent C allele DNA. Digestion with *HaeIII* (indicated in blue) resulted in the undigested amplicon of 201bp in size for the T allele, while fragments of 152bp and 49bp represent C allele DNA. Restriction digest analyses were carried out in a bioanalyzer. Ratios of T and C alleles were calculated from the molar quantities given for the respective fragments greater than 50 bp according to suppliers' instruction for accurate measurement.

Calculation example for RFLP of CD14 rs2569190:

With the given molarities for each digestion fragment, the ratio of CD14 rs2569190 promoter variants was calculated. C = CD14 rs2569190 C allele PCR fragments, T = CD14 rs2569190 T allele PCR fragments, H = hybrid of CD14 rs2569190 C and T allele PCR fragments

Digestion with *HaeIII*:

$$(T + H) : C = 96.2 : 13.4 = 7.18$$

Digestion with *TaaI*:

$$T : (H + C) = 42.1 : 31.9 = 1.32$$

$$T = 100\% - C - H$$

$$T = 56.90\% \quad C = 12.22\% \quad H = 30.88\%$$

$$T = (56.90\% \cdot 2 + 30.88\%) : 2 = 72.34\%$$

$$C = (12.22\% \cdot 2 + 30.88\%) : 2 = 27.66\%$$

$$\text{ratio of CD14 rs2569190 T/C promoter variants} = 2.61$$

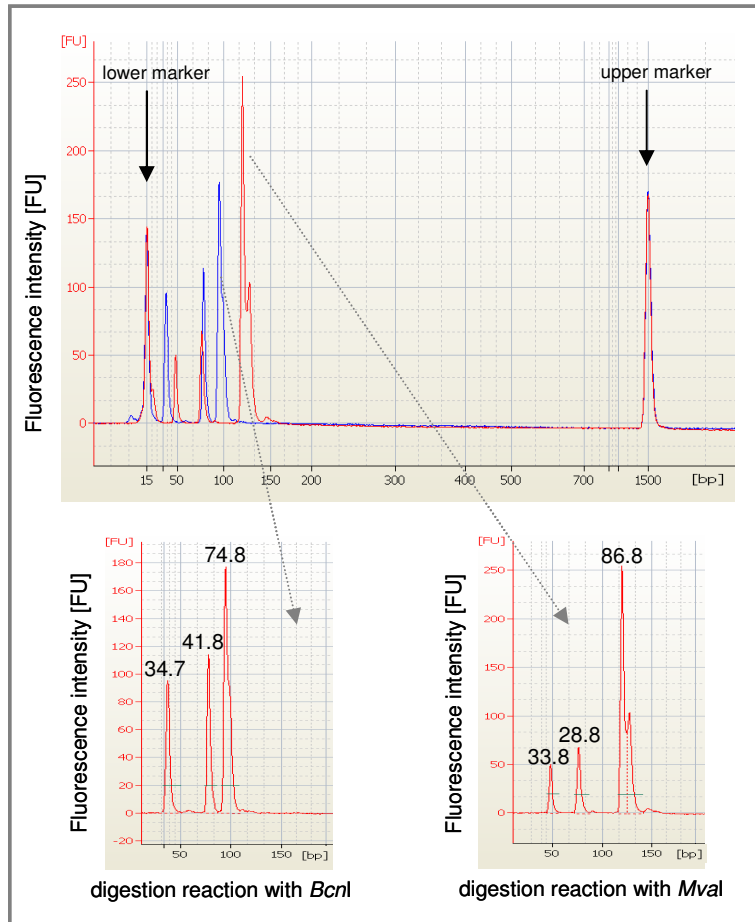


Figure 11b: Analysis of RFLP with the Agilent 2100 bioanalyzer using the example of H5 antibody captured HaploChIP material of individual #1

HaploChIP material (individual #1), was subjected to amplification spanning a region around IRF-1 rs2549009 SNP and yielding a product of 142bp in size. Amplicons were subjected to two different restriction enzyme digestions with *MvaI* and *BcnI*. Digestion with *MvaI* (indicated in red) resulted in fragments of 69bp, 49bp and 24bp in size, representing A allele DNA, while fragments of 118bp and 24bp represent G allele DNA. Digestion with *BcnI* (indicated in blue) resulted in fragments of 91bp, 36bp and 15bp in size representing A allele DNA, while fragments of 72bp, 36bp, 19bp and 15bp represent G allele DNA. Restriction digest analyses were carried out in a bioanalyzer. Ratios of A and G alleles were calculated from the molar quantities given for the respective fragments greater than 50 bp according to suppliers' instruction for accurate measurement.

Calculation example for RFLP of IRF-1 rs2549009:

With the given molarities for each digestion fragment, the ratio of IRF-1 rs2549009 promoter variants was calculated. G = IRF-1 rs2549009 G allele PCR fragments, A = IRF-1 rs2549009 G allele PCR fragments, H = hybrid of IRF-1 rs2549009 G and A allele PCR fragments

Digestion with *MvaI*:

$$(G + H) : A = 86.8 : 28.8 = 3.01$$

Digestion with *BcnI*:

$$G : (H + A) = 41.8 : 74.8 = 0.56$$

$$G = 100\% - A - H$$

$$A = 24.9\% \quad G = 35.9\% \quad H = 39.2\%$$

$$A = (24.9\% \cdot 2 + 39.2\%) : 2 = 45.5\%$$

$$G = (35.9\% \cdot 2 + 39.2\%) : 2 = 55.5\%$$

$$\text{ratio of IRF-1 rs2549009 A/G promoter variants} = 0.82$$

2.4. Sequencing of the CD14 gene

First, three overlapping regions of the CD14 gene and nearby regions spanning 2,101 nucleotides (publication 6.1./Fig. 1) were amplified using two mixtures of 25ng gDNA from 5 different individuals (5ng gDNA per donor) each, heterozygous at CD14 rs2569190, as templates and 0.36 μ M of sequencing primers each (see Fig. 18 and 5.3.). Secondly, the first overlapping region (761 nt) was amplified separately for each of the 10 individuals using 60ng gDNA and sequencing primer pair 1 (5.3.). All products were purified following the spin protocol of the QIAquick PCR Purification Kit (Qiagen) and sequenced by SeqLab (SeqLab, Göttingen, Germany) using an automated sequencer. Alignments and analyses were performed using the BLAST program of NCBI (National Institutes of Health, Bethesda, Md, USA) and version 2.01 of the software Chromas LITE (Technelysium Pty Ltd, Tewantin, QLD, Australia).

2.5. ASTQ

The functional impact of the CD14 and IRF-1 promoter SNPs were investigated in several *in vivo* studies. In addition to HaploChIP analyses, used to investigate the functional relevance of SNPs on promoter activity, the impact of the respective SNPs was determined by ASTQ. The ASTQ technique requires an additional SNP within the transcript as a marker to distinguish transcripts originating from the respective promoter variant. ASTQ of the CD14 and IRF-1 gene was performed in parallel within our research team as part of several medical theses (Aidery, 2010; Bregadze, 2010; Mansur, 2008).

2.6. Haplotyping

To relate HaploChIP results to those obtained by ASTQ, it was necessary to assign allelic transcripts to respective promoter variants. Therefore, haplotyping of different SNPs within the CD14 and IRF-1 gene was carried out on gDNA from individuals heterozygous in both the promoter variant and the ASTQ marker SNP.

Additionally, other genetic variants within the CD14 and IRF-1 gene were also investigated for their linking to respective promoter polymorphisms.

2.6.1. Haplotyping of polymorphisms within the CD14 gene

Haplotyping of 4 genetic polymorphisms within the CD14 gene (rs5744455, rs2569190, rs4914, rs2563298) was performed on gDNA of the larger cohort using different geno-/haplotyping experiments.

In brief, allele-specific primers for CD14 rs5744455 or CD14 rs2569190 were used to generate amplicons of about 2,000 nt in size, spanning all the variant positions of interest. The CD14 rs5744455 or CD14 rs2569190 allelic PCR products were used as templates for distinguishing of the remaining polymorphic positions in different genotyping assays (see 2.2.). Haplotyping was performed in a collaborative approach and also became part of a medical thesis (Bregadze, 2010; Mertens et al., 2009).

2.6.2. Haplotyping of polymorphisms within the IRF-1 gene

In genotyping experiments of the IRF-1 genetic variants rs2549009, rs2549007, rs2706384 and rs839, a correlation between IRF-1 rs2549009 genotypes and genotypes of the other genetic variants was observed. Thus, it was possible to assess haplotyping of these polymorphisms statistically, without further experiments. Therefore, linkage disequilibrium (LD) of haplotypes IRF-1 rs2549009 and rs839 or rs2549007 or rs2706384 was calculated using the software HaploView version 4.1 and expressed as D' and r^2 (Barrett et al., 2005). Unlike D' , which represents the difference between the observed and expected frequencies of haplotypes for two biallelic loci, r^2 indicates the negative relation between LD and the physical distance of sites. Both only have a maximum of 1 if the two loci have the same allele frequencies and if there is a positive association between the most common alleles at the two loci (Awadalla et al., 1999; Hedrick and Kumar, 2001; Lazarus et al., 2002).

2.7. Statistical analyses

Correlation and linear regression analyses were performed using the software GraphPad Prism version 4.00 (GraphPad Software Inc, La Jolla, California, USA). P -values < 0.05 were considered significant. Total IRF-1 gene expression analysis with regard to IRF-1 rs2549009 promoter genotypes was determined using one-way ANOVA (Kruskal-Wallis test) also by applying the software GraphPad Prism version 4.00 (GraphPad Software). Exact test for Hardy-Weinberg equilibrium (HWE) and LD (D' and r^2) were assessed by HaploView version 4.1 (Barrett et al., 2005). HWE is a parameter for the genotypic frequency of two alleles of one gene locus. Calculated on the basis that the alleles A and a have the frequencies p and q ($=1-p$), the equilibrium genotypic frequencies are $AA=p^2$, $Aa=2pq$ and $aa=q^2$ (Mayo, 2008).

3. Results & Discussion

3.1. Establishing and applying HaploChIP to assess the transcriptional activity at CD14 rs2569190 promoter variants in a natural genomic context

Genetic variations within the TLR4 gene have been associated to individual endotoxin sensitivity (Arbour et al., 2000; El-Omar et al., 2008). It is also suggested that genetic polymorphisms within the CD14 gene, in particular CD14 rs2569190 SNP, do determine the outcome of LPS exposure (Baldini et al., 1999; Campos et al., 2005; Levan et al., 2008; Meiler et al., 2005; Virta et al., 2008). Thus, to address the question whether CD14 rs2569190 (C/T) has any functional impact on CD14 promoter activity, HaploChIP analyses using antibodies directed against ser5-P or ser2-P isoforms of the RNAPII were conducted on *ex vivo*-derived PBMC from healthy Caucasians.

Therefore, 13 healthy Caucasian blood donors were first of all genotyped at this variant position and the CD14 rs2569190 heterozygotes (n=9) were used for the investigation of the CD14 promoter activity. Genotype distribution was found to be in line with HWE (Tab. 2). Eleven of the 13 blood donors became part of a larger cohort consisting of altogether 42 healthy Caucasians used for further analyses of the functional impact of CD14 rs2569190 on CD14 transcription in collaboration (see 3.1.4).

Tab. 2: Genotype distribution among 13 healthy Caucasian blood donors

	homozygous CC	heterozygous CT	homozygous TT	MAF	p^a
rs2569190 [C/T]	2 (15.4%)	9 (69.2%)	2 (15.4%)	0.500	0.300

^a p exact test for Hardy-Weinberg equilibrium

According to public databases (NCBI) two different transcript variants of the human CD14 gene exist (Fig. 12). The variant position CD14 rs2569190 which was formerly supposed to be located within the promoter region (at position -159 from transcription start point), is now assumed either to be located within the coding region (transcript variant 1: NM_000591.2) or within intron 1 (transcript variant 2: NM_001040021.1). In experiments designed to determine the splice variant expressed in PBMC, transcript variant 2 was found to be the predominant splice variant, meaning that in PBMC CD14 rs2569190 is located within intron 1 of the CD14 gene (Bregadze, 2010; Mertens et al., 2009).

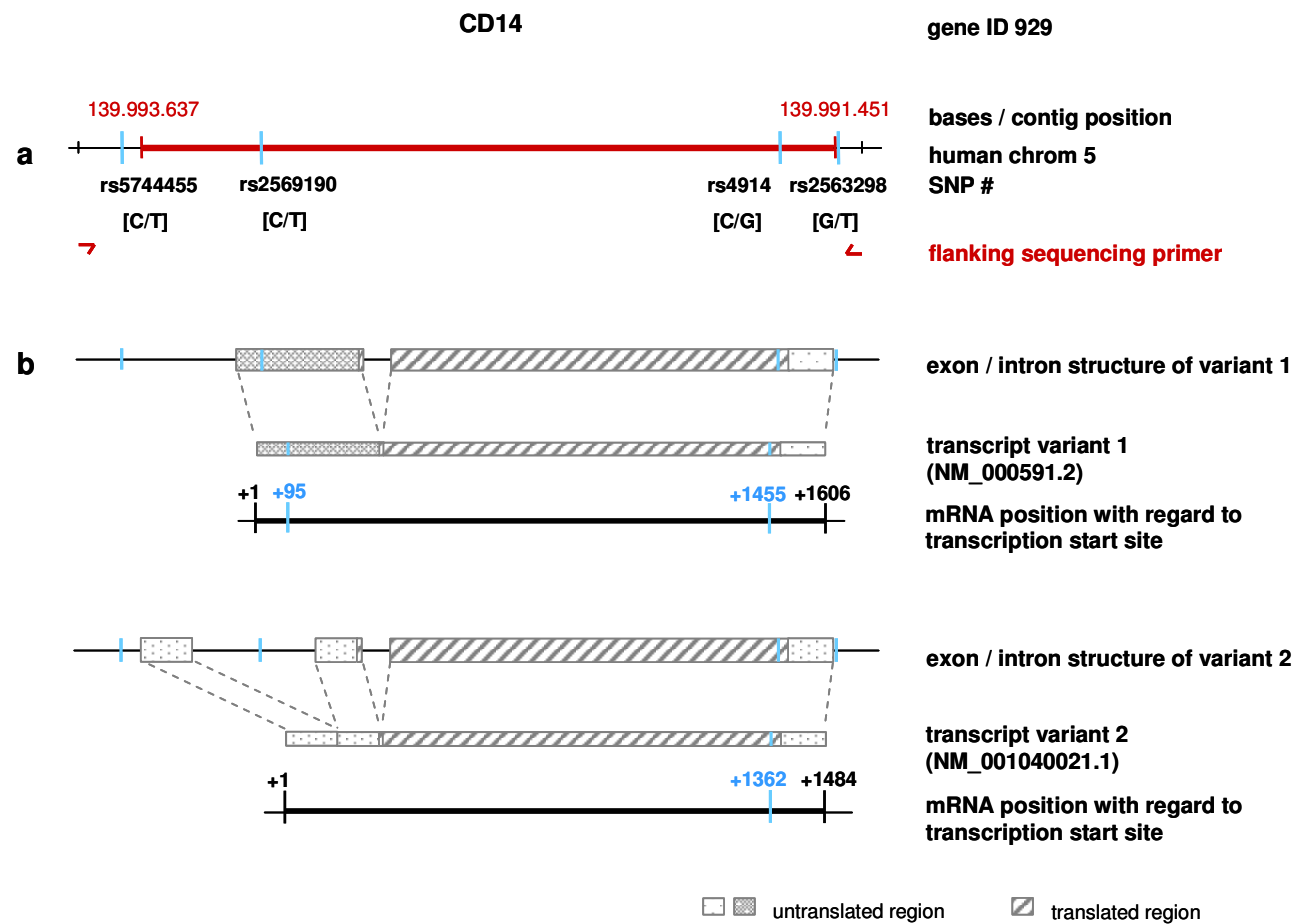


Figure 12: Schematic overview of the organization of the CD14 gene and transcript variant

(a) The CD14 gene is located on chromosome 5. The gene and the nearby 5' and 3' regions contain four SNPs with MAFs > 0.1 as indicated in blue. Presence of these four SNPs was confirmed by sequencing the indicated region. (b) According to public database NCBI, two different transcript variants exist. In transcript variant 1, CD14 rs2569190 is located within the 5' untranslated region (UTR) and because of its intronic location in variant 2, CD14 rs2569190 is absent from transcript variant 2 (Modified from Mertens et al., 2009).

3.1.1. CD14 rs2569190 RFLP analyses with low cycle number PCR

To establish the HaploChIP technique, formalin-fixed and enzymatically sheared chromatin from heterozygous donors was immunoprecipitated by an IgG antibody specific for ser5-P RNAPII. This specifically captured material was subjected to RFLP analysis; including amplification of the immunoprecipitated material, digestion of PCR products with a SNP-specific endonuclease and analysis of digested fragments with a bioanalyzer (see 2.3.3.).

As a measure to avoid the generation of hybrids, low cycle number PCRs with 24-28 cycles were carried out (see 2.3.3.2.). PCR reactions spanning a region around CD14 rs2569190 was performed using antibody-captured material (output). As a control for specific immunoprecipitation, non-captured material (input) was also used as a PCR template. The software NEBcutter V2.0 was used to find a SNP-specific endonuclease that is able to distinguish between the CD14 rs2569190 variant alleles. The enzymes *HaeIII* and *TaaI* were found to be SNP-specific for the CD14 rs2569190 promoter variants. The first enzyme recognizes and digests a DNA sequence containing the C allele, while the second enzyme is specific for a DNA sequence containing the T allele (see 2.3.3.2.). After digestion of PCR products (201bp) with the C allele-specific endonuclease, the amount of digested (CD14 rs2569190 C promoter variant) and undigested PCR products (CD14 rs2569190 T promoter variant) was determined with a bioanalyzer (Fig. 13). Ratios of T and C allele fragments were assessed by quantification of molarities for each fragment given by the 2100 bioanalyzer software. Additionally, as controls for complete and accurate digestion, RFLP analyses were also performed using gDNA from individuals homozygous for the wild-type or variant CD14 rs2569190 alleles. Expectedly, only material derived from CC homozygous individuals showed digested fragments of expected size (152bp and 49bp), while material from TT homozygotes exhibited only undigested PCR products in all samples tested. Input material was found to contain nearly the same amount of C and T promoter variants (T/C ratios 1.09 ± 0.05 - 1.13 ± 0.03). Output material, in contrast, exhibited a higher amount of CD14 rs2569190 T promoter variants (T/C ratios 1.91 ± 0.19 - 2.87 ± 0.04) in all 5 heterozygous donors' samples tested (Mertens et al., 2009).

These findings showed a preferential recruitment of the transcriptionally inactive RNAPII to the CD14 rs2569190 T allele captured by an IgG antibody specific for the ser5-P RNAPII. Thus, this data provides first evidence for a preferential transcription initiation of the variant CD14 rs2569190 allele in *ex vivo*-derived PBMC from healthy Caucasians.

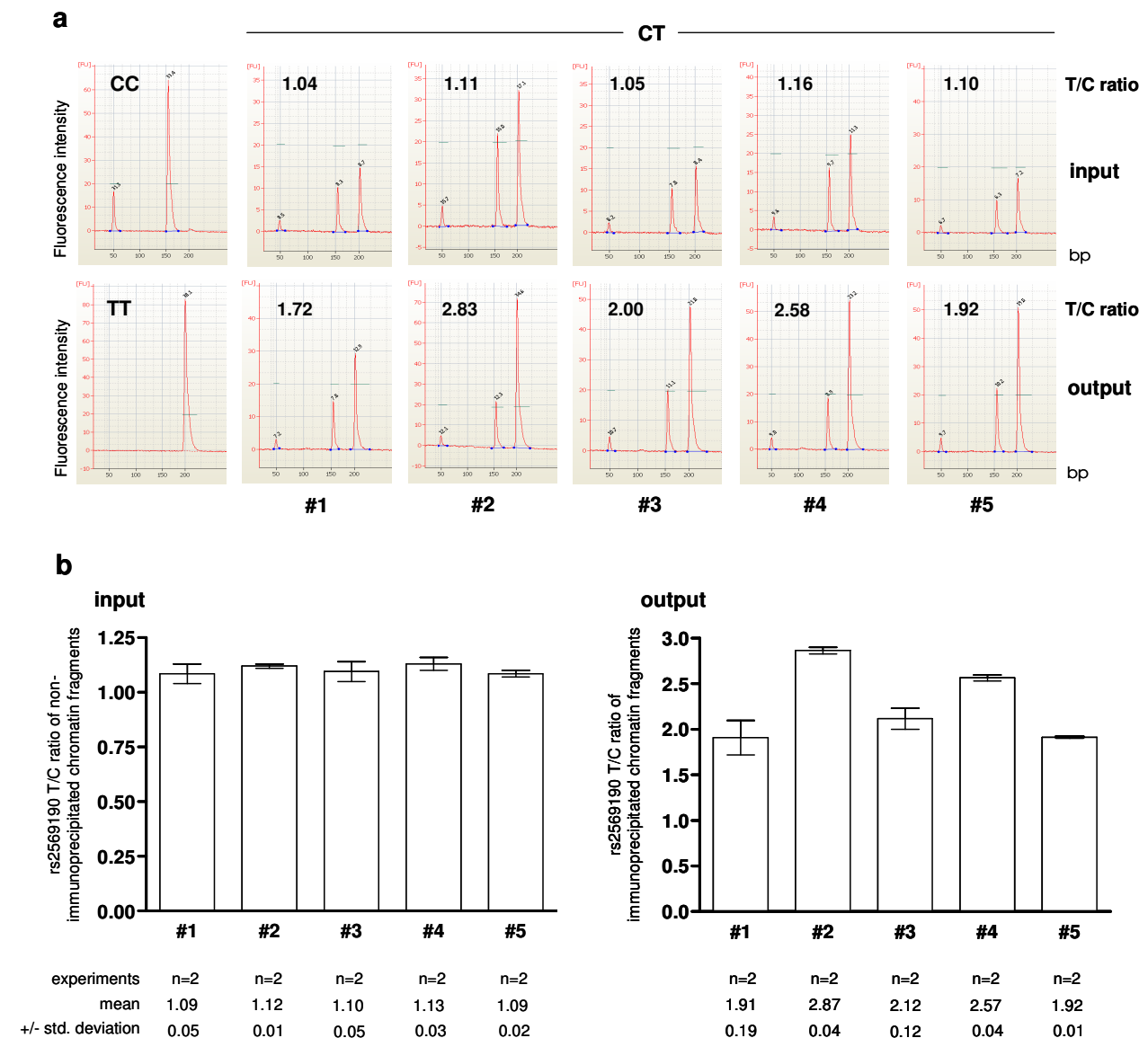


Figure 13: Analysis of the CD14 rs2569190 promoter activity in 5 healthy Caucasian blood donors with HaploChIP

(a) Chromatin from 5 healthy blood donors was immunoprecipitated with an antibody specific for ser5-P RNAPII. Specifically captured and non-captured chromatin was subjected to RFLP analysis, including low cycle number PCRs and digestion reactions with *Hae*III (specific for the CD14 rs2569190 C allele). Materials from digestion reactions were analysed with the 2100 bioanalyzer and T/C ratios were calculated by given molarities for each fragment. Material from homozygous donors (CC and TT) is shown as controls for complete and specific enzymatic activity (Mertens et al., 2009). (b) Experiments were performed twice each, and mean ratios and standard deviations are given for HaploChIP analyses of CD14 rs2569190 heterozygous donors.

As a control, low cycle number PCR RFLP analysis was also performed with the enzyme *TaaI* specific for the T allele in CD14 rs2569190: PCR material from two individuals (#2 and #4) heterozygous in CD14 rs2569190 was subjected to *TaaI* digestion reaction. Analysis for the amount of digested fragments with the 2100 bioanalyzer revealed comparable results: more CD14 rs2569190 T promoter fragments in output material (T/C ratio 2.54; 2.67), while input control showed T/C ratios of ~ 1 (T/C ratio 0.96; 1.06). Controls for accurate enzymatic digestion showed the expected fragments (135bp, 42bp and 24bp) in material from an individual homozygous for the variant T allele in CD14 rs2569190, as well as material from a CC homozygote, which was digested into fragments of the expected 177bp and 24bp (Fig. 14).

Thus, with this additional experiment it was shown first, that RFLP analysis was performed with highly SNP-specific endonucleases and second, that PCR products (input and output material) contained nearly no hybrids that could falsify the generated results.

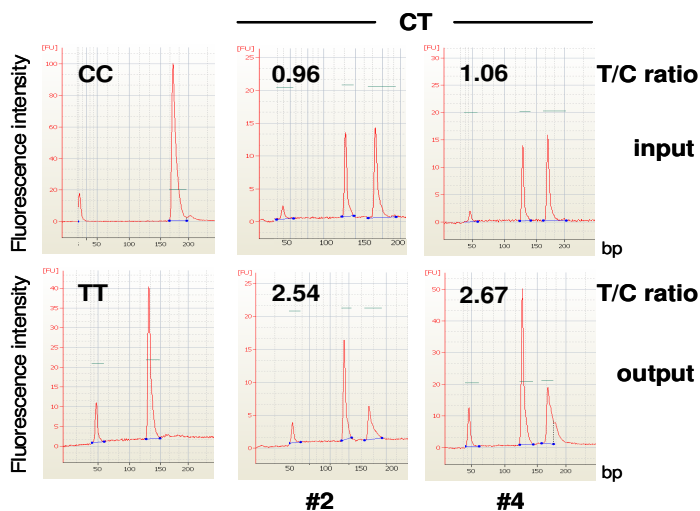


Figure 14: Analysis of HaploChIP material from two individuals (#2 and #4) with RFLP using *TaaI*

Products of low cycle number PCRs from individual #2 and #4 that were also used for digestion with *HaeIII* (Fig. 13) were additionally subjected to digestion reactions with *TaaI*. As controls for accurate enzymatic activity material from homozygous donors (CC and TT) was used.

Because of the continued reproducible results and the expected T/C ratios ~1 for input controls as well as the highly SNP-specific endonucleases the HaploChIP method was found to generate reliable results. But amplification of non- and immunoprecipitated material running low cycle number PCR was proved to be difficult. To avoid the development of hybrids it was necessary to stop PCR within the linear stage of amplification. Therefore, to spot the sufficient cycle numbers of the linear stage, real-time PCRs with input and output material were performed before RFLP. However, because of the different cycling conditions of the different PCR forms, it was not always possible to determine the sufficient cycling numbers.

Thus, analysis of CD14 promoter activity by HaploChIP was conducted on material from 5 donors who were heterozygous in CD14 rs2569190 only.

To obtain reliable results also with conventional PCR, an RFLP approach using two different digestion reactions at a time in combination with a respective calculation was tried.

3.1.2. CD14 rs2569190 RFLP analyses with standard cycle number PCR

Amplicons derived from output material subjected to PCRs of 40-45 cycles were digested by two different SNP-specific enzymes specific for the wild-type allele (*HaeIII*) or the variant allele (*TaaI*) in CD14 rs2569190. Digested fragments were analysed again in the 2100 bioanalyzer. By also taking the generation of hybrids into account, which were assumed to be neither digested by one or the other endonuclease, T/C ratios were calculated (for calculation example see 2.3.3.3.).

As HaploChIP analyses with low cycle number PCRs, output material was found to contain more fragments from the CD14 rs2569190 T variant (T/C ratios 1.44 ± 0.03 - 2.70 ± 0.18) in all 9 heterozygous donors (Fig. 15). Again, controls for specific enzymatic activity showed fragments of expected size. CC homozygous material was digested by *HaeIII* and resulted in fragments of 152bp and 49bp in size, while material from TT homozygotes were left undigested in sizes of 201bp. Digestion reaction with *TaaI* resulted in fragments of 135bp, 42bp and 24bp in size for TT homozygotes and fragments of 177bp and 24bp in size for CC homozygous donors (see 2.3.3.2.).

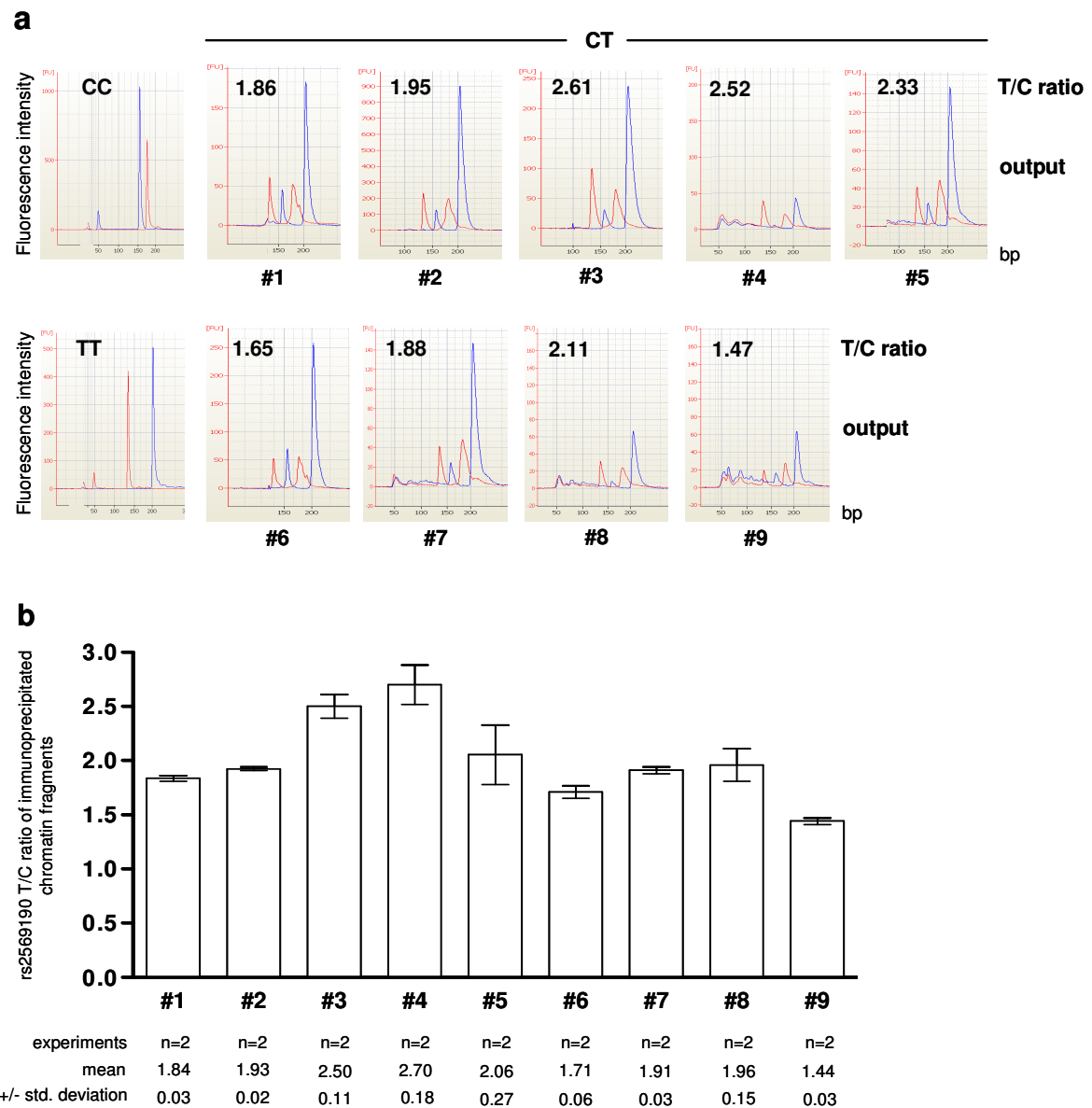


Figure 15: Analysis of the CD14 rs2569190 promoter activity in 9 healthy Caucasian blood donors with HaploChIP

(a) Immunoprecipitated material from 9 heterozygous individuals was subjected to RFLP; including standard cycle number PCR (40–45 cycles) followed by digestion reactions of PCR products with SNP-specific enzymes. Digestion with *HaeIII*, which is specific for the C allele in CD14 rs2569190, is indicated in blue. Digestion with *TaaI* (given in red) is specific for the CD14 rs2569190 T variant (Fragments sizes see 2.3.3.3.). Material from digestion reactions were analysed with the 2100 bioanalyzer. T/C ratios were calculated by given molarities for each fragment and by taking hybrids into account. As controls for correct enzymatic activity, gDNA from homozygous (CC and TT) individuals was additionally subjected to RFLP analysis. (b) Experiments were performed twice each and mean ratios and standard deviations are given.

RFLP analysis, running PCRs with standard cycle numbers followed by digestion reactions with two different SNP-specific enzymes generated reproducible results similar to those achieved with low cycle number PCRs.

These results indicate again a preferential recruitment of the transcriptionally inactive RNAPII to the CD14 rs2569190 T allele in PBMC from healthy Caucasians.

These data on CD14 rs2569190 promoter activity achieved with HaploChIP are in line with *in vitro* experiments suggesting that a DNA element containing the T allele of CD14 rs2569190 favours transcription factor Sp1 and Sp2 binding and therefore promotes gene transcription - at least if the promoter variant is located within a plasmid, i.e. in a non-genomic environment (Baldini et al., 1999; LeVan et al., 2001).

3.1.3. Additional control – immunoprecipitation by a mock isotype (IgG) antibody

Further, to exclude unspecific antibody binding to ser5-P RNAPII, chromatin from 3 heterozygous individuals was additionally immunoprecipitated by a mock isotype (IgG) antibody that should not recognize any human epitope specifically. Input, output and mock-captured material was again analysed via RFLP running standard cycle number PCRs (40-45 cycles), followed by digestion reactions with *HaeIII* and *TaaI*. Digested fragments were analysed again with the 2100 bioanalyzer and T/C ratios were calculated by given molarities for digested and undigested fragments. As expected, input and mock controls showed nearly equal amounts of CD14 rs2569190 T and C allele variant fragments (T/C ratios 1.02-1.21) and output material was found again to exhibit more CD14 rs2569190 T allele fragments (T/C ratios 1.65-1.90). Here as well, controls for correct enzymatic activity were found to contain only fragments of expected size (Mertens et al., 2009). Experiments were repeated later for two individuals and revealed similar results (Fig. 16).

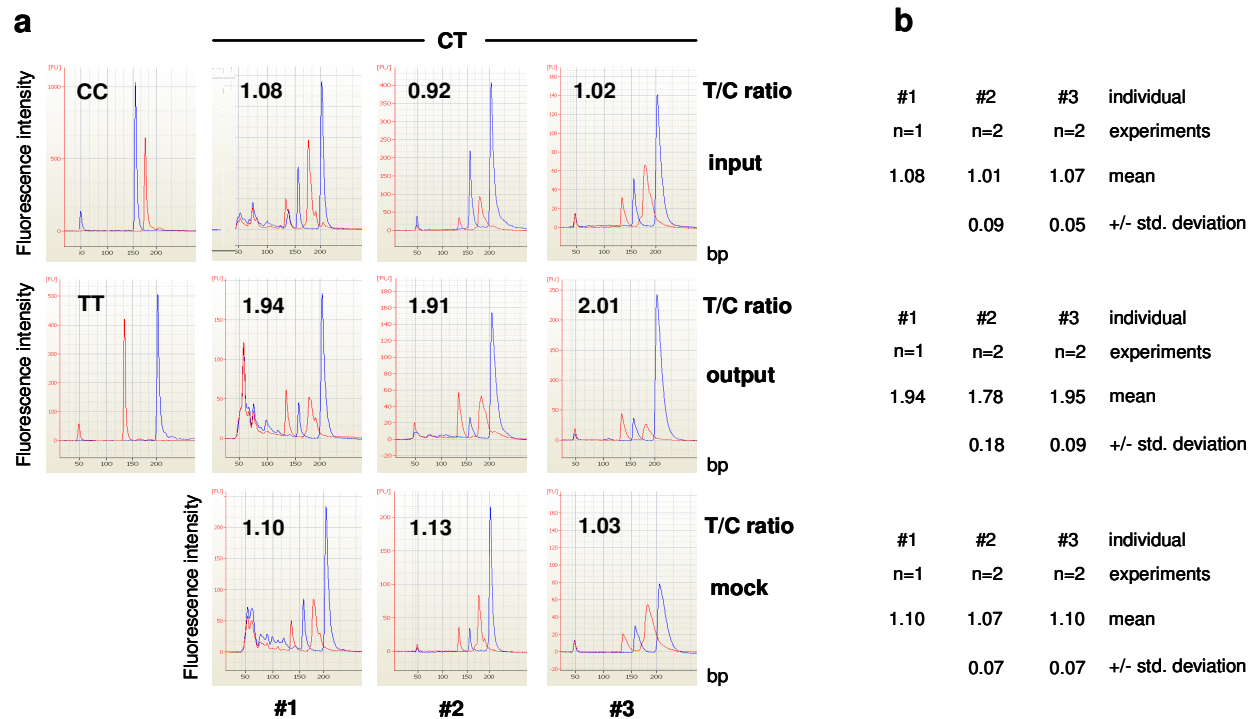


Figure 16: HaploChIP analysis with a mock isotype (IgG) antibody – control for antibody specificity

(a) Chromatin from 3 different individuals heterozygous in CD14 rs2569190 was additionally immunoprecipitated by a mock isotype (IgG) antibody. After amplification of non- and immunoprecipitated materials (input, output and mock control) and digestion reactions with *Hae*III (indicated in blue) and *Tae*I (indicated in red) digested fragments were analysed in the 2100 bioanalyzer (for fragment size see 2.3.3.2.). Ratios of CD14 rs2569190 T and C gene variants were calculated by taking the generation of hybrids into account. As controls for correct enzymatic activity, material from homozygous donors (CC and TT) were also subjected to RFLP analysis. (b) Experiments were performed up to two times und mean ratios and standard deviations are given.

By additional immunoprecipitation with a mock isotype antibody, it was shown that output material was specifically captured (by an anti-ser5-P RNAPII antibody). These results of control materials confirm the assumption that the HaploChIP technique is a reliable tool for investigation of the activity of different CD14 promoter variants by assessing allelic transcription initiation.

3.1.4. Lack of correlation between CD14 rs2569190 promoter activity and allelic CD14 transcription in PBMC

Investigations of the CD14 rs2569190 promoter activity with the HaploChIP technique (both with low and standard cycle number PCRs) showed a preferential binding of the ser5-P RNAPII to the T allele of CD14 rs2569190 in PBMC from healthy individuals. Thus, it was expected that more CD14 transcripts would be found derived from the CD14 rs2569190 T allele than from the wild-type allele in PBMC.

Allele-specific CD14 transcription was analysed using the SNP rs4914 which is located within the CD14 transcript as a marker to distinguish CD14 transcripts derived from different promoter variants (Fig. 12). ASTQ was thus applied to material from individuals heterozygous both in CD14 rs2569190 and in CD14 rs4914. To assign allelic CD14 rs4914 transcripts to respective CD14 rs2569190 promoter variants, haplotyping was performed (see publication 6.1). However, no excess of allelic CD14 transcripts was detected in PBMC from 7 double-heterozygous (CD14 rs2569190 and CD14 rs4914) donors (Bregadze, 2010; Mertens et al., 2009) (donors #6 and #7 were used for both analyses, HaploChIP and ASTQ).

To address the question of whether another SNP than CD14 rs2569190 might influence CD14 transcription, genotyping of variant positions within the 5' and 3' region of the CD14 gene with MAFs >0.1 (rs5744455 - rs2569190 - rs4914 - rs2563298) as well as geno-/haplotyping were performed using a combination of ARMS-PCR and genotyping assays (Fig. 12) in collaboration in our research team on material from a total of 42 healthy Caucasians (see publication 6.1).

Genotype distribution of these 4 SNPs was found to be in line with HWE and MAFs (0.26 - 0.44 - 0.15 - 0.27) were found to be similar to published MAFs for Caucasians at NCBI (0.30 - 0.47 - 0.17 - 0.21). Additionally, total CD14 transcripts were found to be neither related to genotypes of the four variant positions of interest nor to C-T-C-G haplotype (the only one carrying the T allele of CD14 rs2569190) or non-C-T-C-G haplotypes in PBMC (Bregadze, 2010; Mertens et al., 2009). Taken together, no correlation between allele-specific CD14 promoter activity and gene transcription was found in PBMC of healthy Caucasians.

3.1.5. Validation of HaploChIP and ASTQ results using the imprinted gene SNRPN

Because of a lack of correlation between CD14 rs2569190 promoter activity and allele-specific transcripts or CD14 transcription rate, control experiments for both methods - HaploChIP and ASTQ - were performed. Therefore, input and output material as well as genomic and complementary DNA (cDNA) were genotyped at the SNP SNRPN rs705 (G/A) within the transcript of the imprinted gene SNRPN. Control experiments were carried out in real-time PCRs using allele-specific probes for discrimination of the SNRPN rs705 variant alleles. While G and A variants were expected to be found in input and gDNA, output material and cDNA were expected to contain only one SNRPN rs705 variant because of monoallelic expression of the imprinted gene.

For validation of HaploChIP results, non- and immunoprecipitated material from 10 blood donors was available. In the input material, one donor was found to be homozygous for the SNRPN rs705 wild-type allele G, while 4 individuals were found to be homozygous for the variant allele A and 5 donors were found to carry both variant alleles. All SNRPN rs705 heterozygotes were found to exhibit only the G or the A allele in output material, indicating a selective binding of the RNAPII to only one variant allele of SNRPN rs705, as expected for monoallelic expression (Fig. 17a).

Validation of ASTQ experiments was conducted on material from 42 blood donors. In gDNA, 8 blood donors exhibited the SNRPN rs705 homozygous GG genotype, while 17 individuals each were found to carry only the allele A or both variant alleles. As expected, all SNRPN rs705 heterozygotes who carried both variant alleles in gDNA, showed only the G or the A allele expressed in cDNA, indicating here as well monoallelic expression of the imprinted gene SNRPN (Mertens et al., 2009).

Additionally, all individuals (#a - #e) found to be heterozygous at SNRPN rs705 in input material contained not only one, but also the same variant allele in output material and in mature mRNA transcripts as well (Fig. 17b). Hence selective binding of the ser5-P RNAPII to only one variant allele of SNRPN rs705 seems to result in monoallelic expression of the preferred allele, as expected.

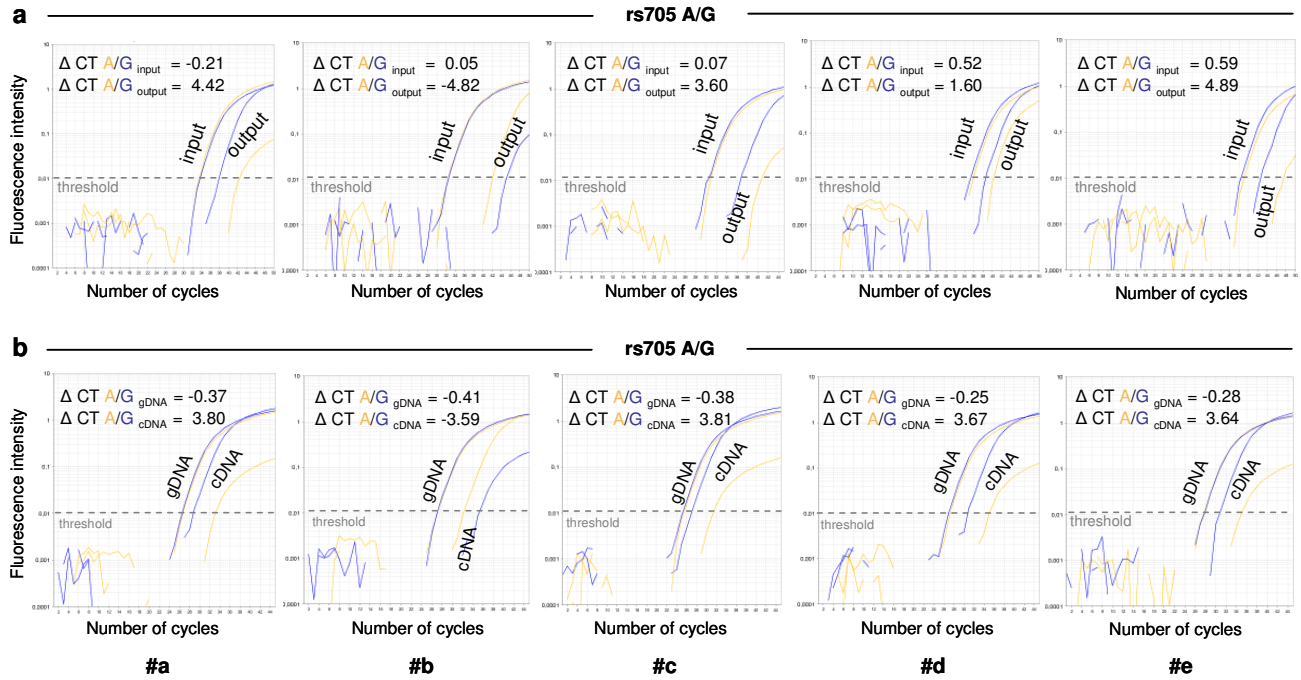


Figure 17: Validation of HaploChIP and ASTQ results

For validation of (a) HaploChIP and (b) ASTQ results, input and output material as well as genomic and cDNA were investigated for the presence of variant alleles at SNRPN rs705 (G/A) located within the transcript of the imprinted gene SNRPN. ΔCT values for SNRPN rs705 G (blue) and A (orange) alleles in output material and cDNA were significantly different to respective input and gDNA values, as expected for monoallelic gene expression.

With the expected monoallelic expression of the imprinted gene SNRPN in both output and cDNA, results from HaploChIP and ASTQ experiments could be confirmed. Thus, the non-consistent HaploChIP and ASTQ results might be due to the impact of a yet - undetected polymorphism or haplotype (see 3.1.6.) - or to other reasons (see 3.1.7.).

3.1.6. Sequencing of the CD14 gene and flanking 5' and 3' prime region

Furthermore, to address the question of whether any undetected SNP within the CD14 gene might influence ser5-P RNAPII binding to the promoter, PCR products of the CD14 gene and flanking 5' and 3' region from 10 individuals heterozygous in CD14 rs2569190 (partly taken from the larger cohort of 42 healthy Caucasians) were sequenced (Fig. 24) (see 2.4). First, two mixtures of CD14 PCR products from each of 5 heterozygous individuals were sequenced. Both mixtures showed at least two possible unknown variant positions within the promoter region (Fig. 18a). Next, sequencing of single amplification products of the 5' region of the CD14 gene from every individual revealed no additional variant position (Fig. 18b) (Mertens et al., 2009).

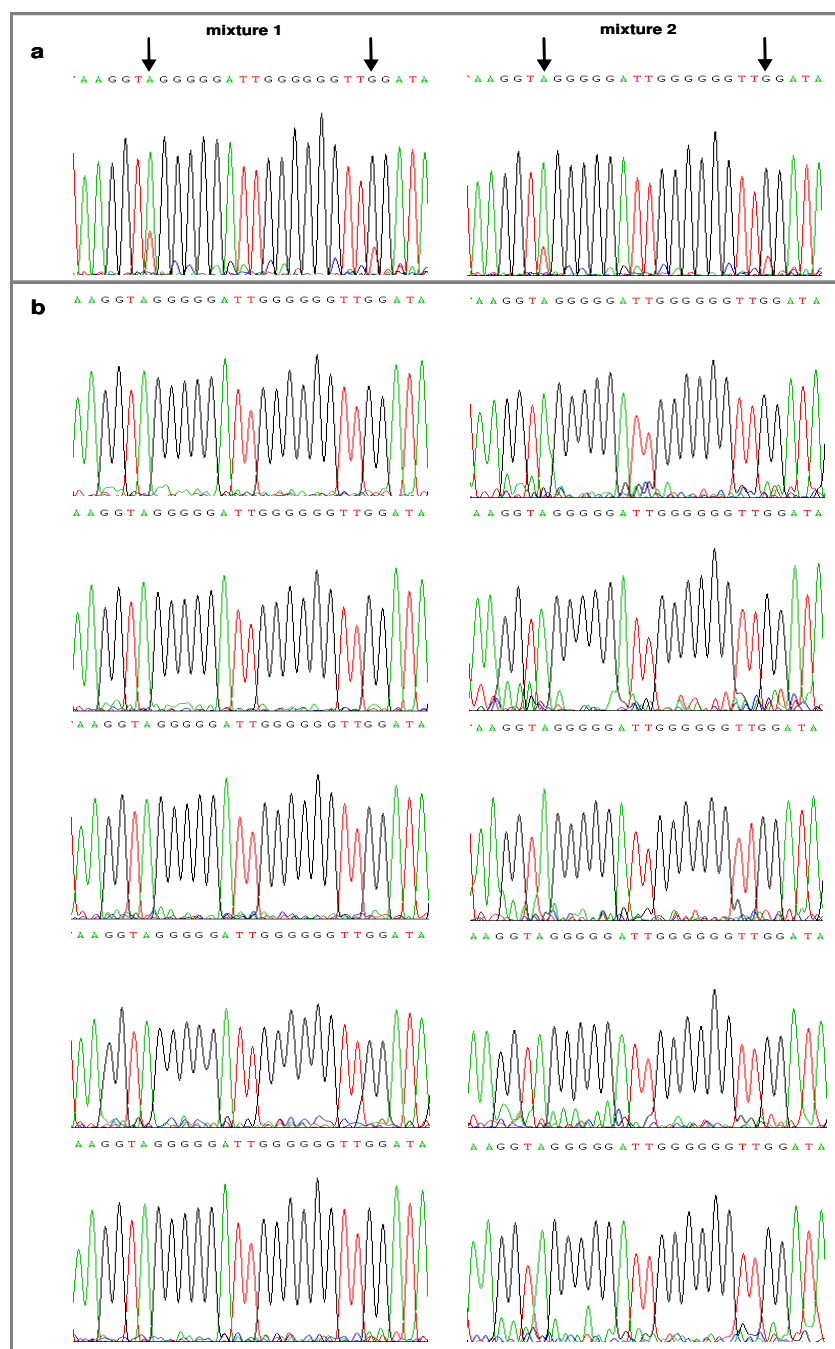


Figure 18: Sequencing of the CD14 gene and nearby regions

Cut-outs of the sequenced promoter region of the CD14 gene are given. (a) Sequence of the 5' region of CD14 from two mixtures of PCR products from each of 5 different individuals heterozygous in CD14 rs2569190 is shown. Two possible unknown variant positions located at positions -103 and -86 (from transcription start point) were detected and are indicated by arrows. (b) Sequence analysis of the 5' region of the CD14 gene for each of the 10 single individuals revealed no unknown SNPs.

Cut-outs were taken from analyses of sequencing data performed with the software Chromas LITE.

3.1.7. HaploChIP analysis with antibodies specific for ser5-P/ser2-P RNAPII - CD14 rs2569190 allele-specific initiation of transcription but absence of allelic elongation in PBMC

PBMC were found to contain CD14 rs2569190 within intron 1, indicating that this SNP might not have any influence on mRNA stability, at least in PBMC (Bregadze, 2010; Mertens et al., 2009). A more likely effect of this SNP on transcription factor binding can be hypothesised. It has been shown in reporter gene assays and EMSA by LeVan et al. that the T allele benefits transcriptional activity in monocytic cells by an enhanced affinity to the activating transcription factors Sp1 and Sp2 (LeVan et al., 2001). *In vivo*, this effect might be seen in a stronger binding of ser5-P RNAPII to T allele variants and thus a stronger transcription initiation of the CD14 rs2569190 T allele in PBMC.

Preferential recruitment of ser5-P RNAPII does not result in a higher transcription of the preferred allele. As mentioned before, it is known that phosphorylation of CTD at the largest subunit of RNAPII is a key event during mRNA metabolism. The phosphorylation at serine 2 is essential for transcription elongation (Palancade and Bensaude, 2003). It seems that the very preferential recruitment of RNAPII, phosphorylated on serines at position 5, to CD14 rs2569190 T allele variants does not necessarily lead to fully phosphorylation on serines at position 2 and thus does not result in a higher transcription rate of the preferred allele. It was shown that the ser5-P RNAPII pre-assembles to gene promoters in mammals, but however, does not generate mature mRNA transcripts. The control of gene expression was shown to depend on the transition from transcription initiation to elongation by phosphorylation at serines 2 at the RNAPII (Hargreaves et al., 2009).

To determine whether the CD14 polymorphism rs2569190 has any effect on ser2-P RNAPII and therefore on allele-specific transcription elongation, HaploChIP analyses were performed using H5 and H14 (IgM) antibodies as well as a mock isotype (IgM) antibody. Therefore, an additional bridging step before precipitation of chromatin was necessary (see 2.3.2.). HaploChIP material was then subjected to RFLP running standard cycle number PCRs followed again by digestion reactions with *HaeIII* and *TaaI* and analysis with the 2100 bionalyzer. Therefore, input and immunoprecipitated material of three different individuals, one (#2) only heterozygous in CD14 rs2569190 and two (#6 and #7) double heterozygous (CD14 rs2569190, CD14 rs4914), was available (Fig. 19).

As expected, control material (input and mock control) contained nearly the same amount of CD14 rs2569190 T and C gene variants (T/C ratio 0.89-1.16) and H14-immunoprecipitated material showed again about twice the number of CD14 rs2569190 T variant fragments (T/C ratio 1.74-1.94), also indicating in this experiment that there is a preferential binding of the ser5-P RNAPII.

In contrast, H5-captured chromatin exhibited similar amounts of CD14 rs2569190 T and C gene variants (T/C ratio 0.91-1.01) as did the mock control.

More precisely, nearly no difference in the amount of CD14 rs2569190 variant fragments was detected in specific- (H5) and unspecific- (mock) captured material. These results indicate an absence of a preferential recruitment of the transcriptionally active RNAPII to the CD14 rs2569190 variants or the respective CD14 haplotypes. These results reflect the findings on allelic CD14 transcription in PBMC, at least as evidenced in two double heterozygous individuals (#6, #7; also same individuals as in (Mertens et al., 2009)).

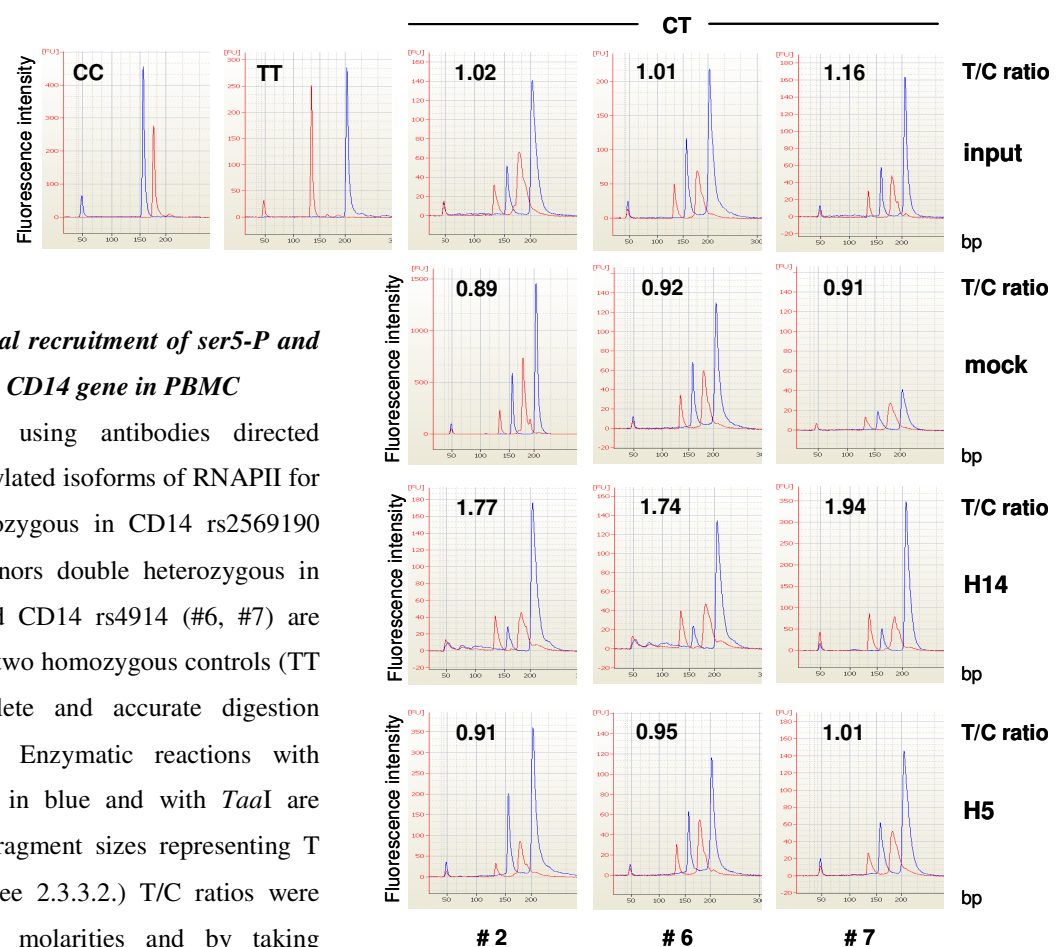


Figure 19: Differential recruitment of ser5-P and ser2-P RNAPII to the CD14 gene in PBMC

HaploChIP analyses using antibodies directed against two phosphorylated isoforms of RNAPII for one individual heterozygous in CD14 rs2569190 (#2) and for two donors double heterozygous in CD14 rs2569190 and CD14 rs4914 (#6, #7) are shown. Additionally, two homozygous controls (TT and CC) for complete and accurate digestion reactions are given. Enzymatic reactions with *HaeIII* are indicated in blue and with *TaaI* are marked in red (for fragment sizes representing T and C allele DNA see 2.3.3.2.) T/C ratios were calculated by given molarities and by taking hybrids into account. Complete experiments were performed once for each tested individual, so far.

It was shown that HaploChIP analysis with the ser5-P H14 antibody confirmed the results generated by an IgG antibody specific for ser5-P RNAPII: a higher amount of CD14 rs2569190 T gene variants in immunoprecipitated chromatin from heterozygous individuals. But H5-captured material, on the contrary, showed no accumulation of specific fragments, nor did controls (input and mock material). Taken together, with expected T/C ratios ~1 for input and mock controls, highly SNP-specific restriction enzymes for the CD14 polymorphism rs2569190 and antibodies that are able to distinguish between the inactive and active form of the RNAPII the HaploChIP method was established as a successful tool for investigating the activity of different CD14 promoter variants, assessing both allelic transcription initiation and elongation in PBMC.

Due to the intronic location of CD14 rs2549190 in PBMC from the healthy cohort, an effect on mRNA stability can be ruled out. As mentioned before, an influence on transcription regulation in terms of an enhanced affinity to transcription factor binding, shown by *in vitro* experiments (LeVan et al., 2001), is more likely. But also an effect of CD14 rs2569190 on an intron-mediated regulation of CD14 gene expression is possible. It is known that introns can play a role in eukaryotic gene expression by harbouring enhancer elements or alternative promoters (Rose, 2008).

But based on the fact that the T allele of CD14 rs2569190, or the C-T-C-G haplotype, seems only to be advantageous for transcription initiation but not for transcription elongation, and the lack of an allelic CD14 gene expression imbalance in PBMC (Bregadze, 2010; Mertens et al., 2009), endotoxin sensing appears not to rely on a functional relevance of CD14 rs2569190 in PBMC.

Regulatory variants can have inconsistent effects in different tissues (Pastinen and Hudson, 2004). It has been shown that CD14 is differentially expressed in monocytes and hepatocytes and that the latter contribute to sCD14 production (Pan et al., 2000). Thus, it appears more likely that the association of CD14 rs2569190 with higher sCD14 levels and lower IgE levels is due to an effect of the SNP in other cells lines, like non-myeloid cells, e.g. liver cells (Mertens et al., 2009).

3.2. Establishing and applying HaploChIP to assess the transcriptional activity of IRF-1 rs2549009 promoter variants in a natural genomic context

According to public databases, the IRF-1 gene shows a high genetic variability in various ethnic populations. For the Caucasian population there are 25 variant positions with a MAF >0.1 within the IRF-1 gene and 5' and 3' region known so far (NCBI). Some IRF-1 polymorphisms have been associated with various disease conditions in different ethnic populations. For example, intronic variations have been identified as linked to an altered susceptibility to human immunodeficiency virus 1 (HIV-1) infection in a Kenyan cohort (Ball et al., 2007). Further, a different set of polymorphisms within the IRF-1 gene, amongst others the IRF-1 rs2549009 (G/A) SNP within the promoter region, was found to be associated with HCV infection in Caucasians (Fortunato et al., 2008; Saito et al., 2002; Wietzke-Braun et al., 2006) (Fig. 20). Thus, to address the question of whether the variant position IRF-1 rs2549009, located at position -300 from the transcription starting point, confers any functional impact on IRF-1 promoter activity, HaploChIP analyses were performed on *ex vivo*-derived PBMC from healthy Caucasians.

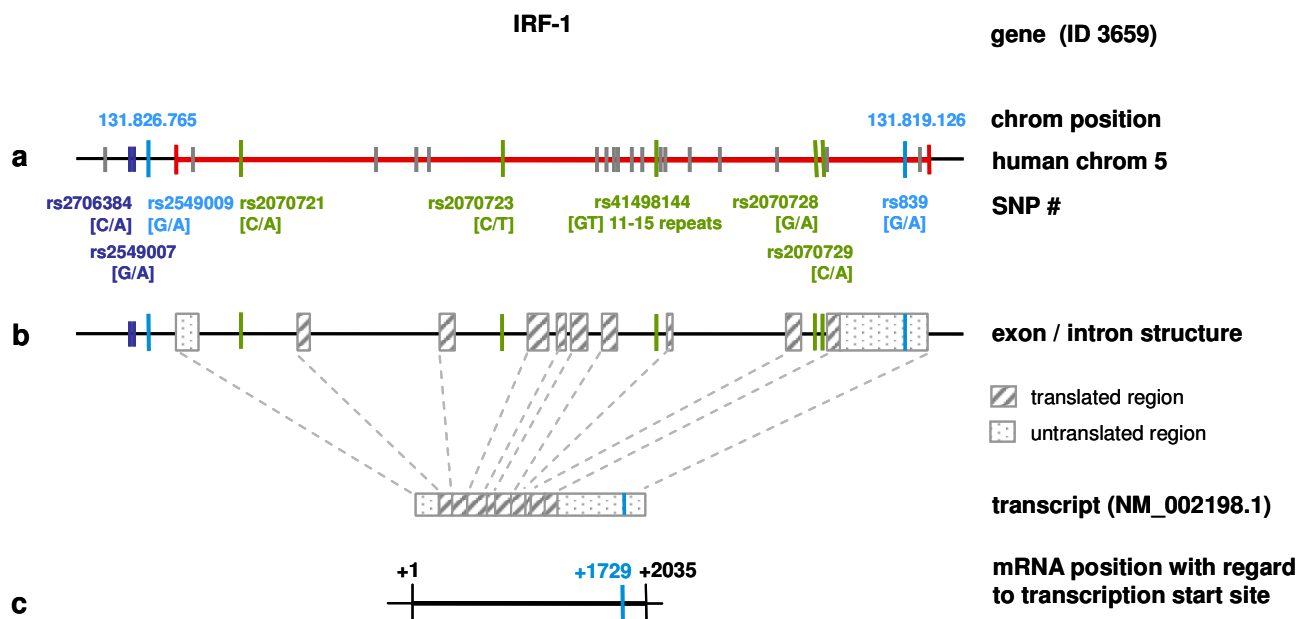


Figure 20: Schematic overview of the organization of the IRF-1 gene und transcript

In humans, the IRF-1 gene is located on chromosome 5. The gene and the nearby 5' and 3' region include 25 variant positions with a MAF > 0.1 known so far for the Caucasian population (NCBI). The variant position IRF-1 rs2549009 and the ASTQ marker SNP IRF-1 rs839 are indicated in light blue within the chromosome (a), the gene (b), and the transcript (c). Two other SNPs at position -410 (IRF-1 rs2549007) and -415 (IRF-1 rs2706384) that have been described as in a strong LD with IRF-1 rs2549009 (Saito et al., 2001) are marked in dark blue. Intronic variations that have been described as associated with different virus diseases are given in green (Ball et al., 2007; Fortunato et al., 2008) (Modified from Mertens et al., 2010).

Therefore, 13 healthy Caucasian blood donors were first of all genotyped at this variant position and IRF-1 rs2549009 heterozygous individuals (n=6) were used for the investigation of the allelic IRF-1 promoter activity. Genotype distribution was found to be in line with HWE (Tab. 3). Eleven of these 13 blood donors are part of a larger cohort of healthy Caucasians with available material from 41 donors that was used for further functional analyses of the transcriptional activity of the IRF-1 promoter variants (see 3.2.3).

Tab. 3: Genotype distribution among 13 healthy Caucasian blood donors

	homozygous GG	heterozygous GA	homozygous AA	MAF	p^a
rs2549009 [G/A]	6 (46.2%)	6 (46.2%)	1 (7.6%)	0.308	1.00

^a p exact test for Hardy-Weinberg equilibrium

3.2.1. IRF-1 rs2549009 RFLP analyses with standard cycle number PCR

Here, the HaploChIP technique was established and performed on formalin-fixed and enzymatically sheared chromatin available from 5 healthy IRF-1 rs2549009 heterozygotes. The question of whether IRF-1 rs2549009 has any functional impact on IRF-1 promoter activity was assessed by investigating the relative binding frequency of differentially phosphorylated RNAPII isoforms to IRF-1 promoter variants. (i) HaploChIP material was immunoprecipitated by the IgG antibody that is specific for ser5-P RNAPII. (ii) HaploChIP analysis was also performed by immunoprecipitation using the IgM antibodies that distinguish between transcriptionally inactive (H14) and active (H5) RNAPII and the mock isotype (IgM). For IgM antibodies, additional bridging between protein G magnetic beads and an anti-IgM-specific IgG antibody was necessary before precipitation of chromatin (see 2.3.2.). Antibody-captured and non-immunoprecipitated (input) chromatin was subjected to RFLP; this process included amplification (40-45 cycles) spanning a region around IRF-1 rs2549009, followed by digestion reactions of PCR products and analysis of digested material with the 2100 bioanalyzer. The software NEBcutter V 2.0 was again used to identify SNP-specific endonucleases.

*Mva*I, which recognizes and digests a DNA sequence containing IRF-1 rs2549009 A, was found, while *Bcn*I was observed to be specific for a DNA sequence containing the IRF-1 rs2549009 wild-type G (see 2.3.3.2.). For control of accurate digestion reactions, RFLP analyses were additionally performed using gDNA from individuals homozygous for the IRF-1 rs2549009 wild-type and the variant allele.

(i) Input material contained, as expected, nearly the same amount of IRF-1 rs2549009 A and G promoter variants (A/G ratios $1.00 \pm 0.01 - 1.11 \pm 0.03$). But output material, that was immunoprecipitated by the IgG antibody directed against ser5-P RNAPII, exhibited a higher amount of IRF-1 rs2549009 A promoter variants (A/G ratios $1.24 \pm 0.01 - 2.11 \pm 0.31$) in all samples tested. Controls for digestion reactions showed fragments of expected size (Fig. 21).

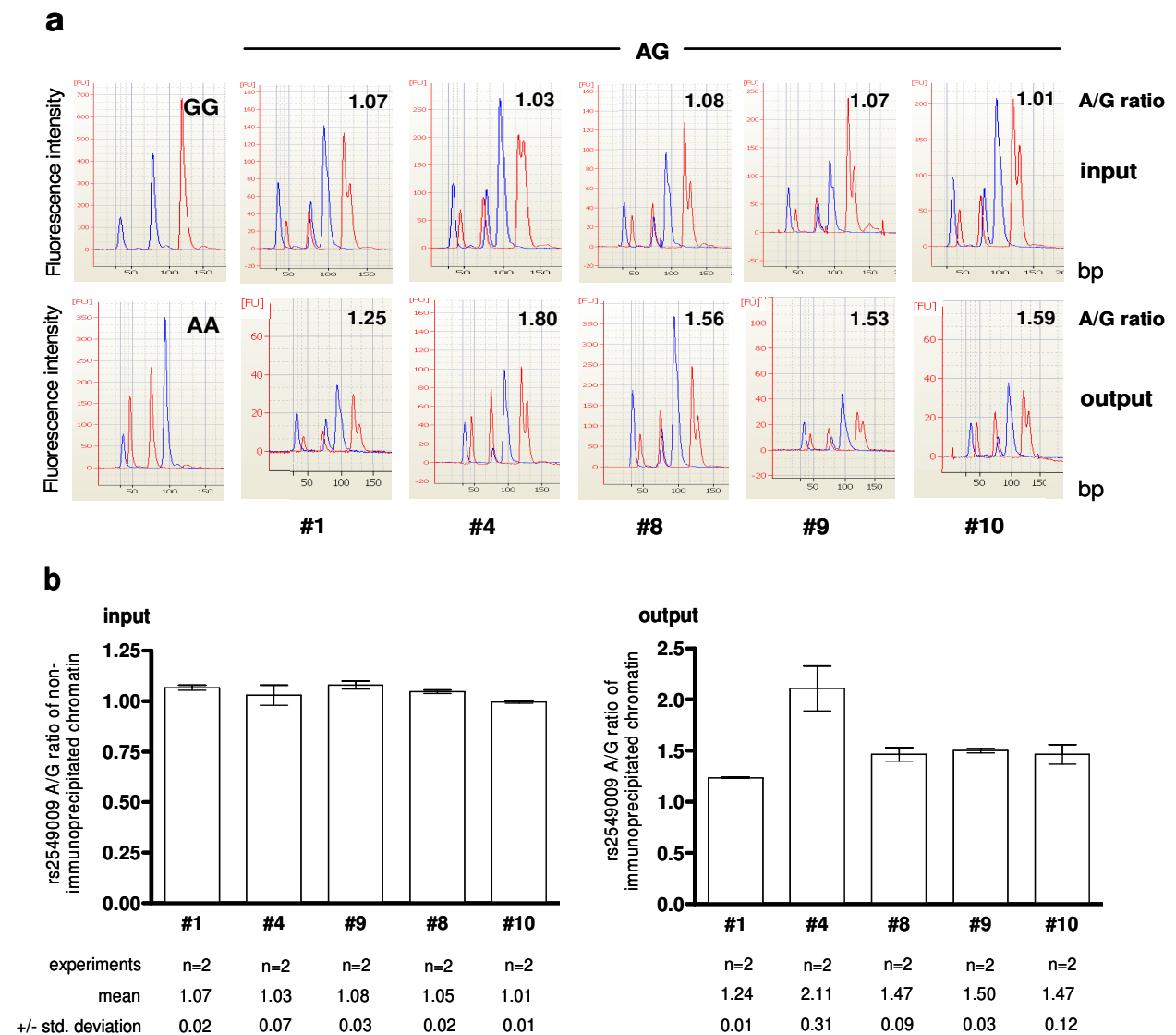


Figure 21: HaploChIP analysis of the IRF-1 rs2549009 promoter activity

(a) Formalin-fixed and enzymatically sheared chromatin from 5 healthy individuals, heterozygous in SNP IRF-1 rs2549009, was immunoprecipitated with an antibody directed against ser5-P RNAPII. Non- and immunoprecipitated chromatin was subjected to amplification spanning a region around IRF-1 rs2549009. Products of 142bp in size were digested by two different restriction enzymes, *MvaI* and *BcnI*. Digestion with *MvaI* is indicated in red, while digestion with *BcnI* is shown in blue (Fragment sizes see 2.3.3.2.). Digestion analyses were carried out in a bioanalyzer. Material from homozygous individuals (AA and GG) is shown as controls for complete and specific enzymatic activity. (b) Experiments were performed twice; mean ratios and standard deviations are given for each individual.

(ii) Controls for accurate digestion with *MvaI* and *BcnI* showed digested fragments of expected size. Input and mock (IgM) control-captured material showed, as expected, almost no difference in IRF-1 rs2549009 A and G promoter variants (A/G ratio 0.98 – 1.08). In all H14-captured material, the transcriptionally inactive ser5-P RNAPII was again found to preferentially bind the IRF-1 rs2549009 A allele (A/G ratio 1.22 ± 0.08 – 1.40 ± 0.05) (see publication 6.2., page 71; Mertens et al., 2010). In contrast, the ratio of A and G promoter variants in H5-captured material was found to vary between 0.79 ± 0.01 (individual #1) and 1.24 ± 0.02 (individual #9). Some donors (#8; #10) showed almost no difference (A/G ratio 0.96 ± 0.04 ; 1.05 ± 0.05) in the amount of IRF-1 rs2549009 A and G promoter fragments, while other individuals were found to exhibit a higher amount of the IRF-1 rs2549009 G or A promoter variant (Fig. 22).

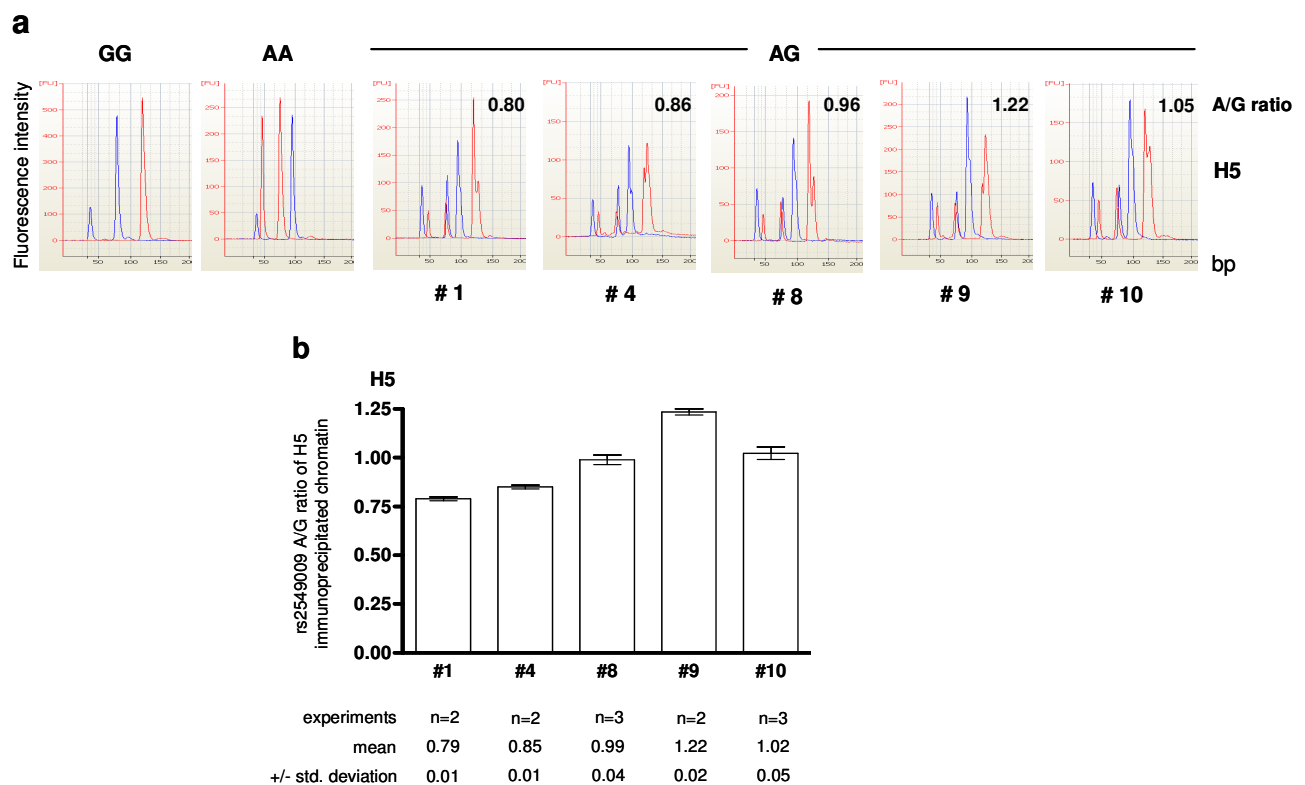


Figure 22: HaploChIP analysis of the IRF-1 rs2549009 promoter activity: H5 antibody specific for ser2-P RNAPII

(a) Immunoprecipitation by H5 antibody was conducted on chromatin from 5 healthy donors heterozygous in IRF-1 rs2549009. HaploChIP material was subjected to RFLP; this included amplification spanning a region around IRF-1 rs2549009 and digestion of PCR products. Digestion reactions were performed using *MvaI* (red) and *BcnI* (blue), followed by analysis of fragments in a bioanalyzer. IRF-1 rs2549009 A/G ratios were calculated by given molarities for each fragment by the 2100 bioanalyzer software. Accurate digestion reactions were controlled by material from IRF-1 rs2549009 homozygous individuals (AA and GG) that was additionally subjected to RFLP. For fragment sizes see 2.3.3.2. (b) Experiments were performed at least twice. Mean ratios and standard deviations are given (Modified from Mertens et al., 2010).

As expected, input and mock controls always showed A/G ratios of ~1, as did also the controls (AA and GG material) for accurate and specific digestion reactions, which exhibited only, as expected, fragments of expected size in all samples tested. These controls showed that immunoprecipitation by H14, H5 and IgG antibodies, despite the additional bridging step for H14 and H5, was specific. Additionally, RFLP analysis was performed using highly IRF-1 rs2549009 SNP-specific endonucleases, *MvaI* and *BcnI*. Thus, the HaploChIP method was established successfully as a tool to investigate the activity of the IRF-1 rs2549009 promoter variants by distinguishing phosphorylated isoforms of the RNAPII during the transcription cycle.

The HaploChIP results show a preferential binding of the transcriptionally inactive RNAPII to IRF-1 rs2549009 A promoter variant, but no preferential recruitment of the ser2-P enzyme to the one or the other IRF-1 rs2549009 promoter variant. This indicates that the IRF-1 rs2549009 A promoter variant may be advantageous for transcription initiation, but an allelic IRF-1 transcription elongation seems not to be solely controlled by this promoter polymorphism in PBMC (Mertens et al., 2010). The observation of a donor-dependent allele-specific ser2-P RNAPII activity is compatible with *in vitro* findings on IRF-1 promoter activity showing stronger promoter activity either of the one or the other allele, depending on the system. Saito et al. showed a stronger induction of gene expression for the IRF-1 rs2549009 variant allele A in unstimulated or IFN-stimulated PLC/PRF/5 hepatoma cells, while less activity of a gene construct carrying the IRF-1 rs2549009 A allele was found in ionomycin-stimulated Jurkat T cells by Schedel and colleagues (Saito et al., 2001; Schedel et al., 2008).

3.2.2 Correlation between allele-specific IRF-1 rs2549009 promoter activity and allelic transcription of IRF-1 in PBMC

A donor-dependent preferential recruitment of the ser2-P RNAPII to the IRF-1 rs2549009 promoter variants was shown with HaploChIP experiments, indicating differential allelic IRF-1 gene transcription elongation. Thus, it was also expected to find an allele-specific imbalance of gene expression in IRF-1 transcripts. The relative amount of allelic transcripts was therefore quantified using IRF-1 rs839, located within IRF-1 exon 10, as a marker SNP for ASTQ (Fig. 20). Assessing allelic IRF-1 transcripts to respective IRF-1 rs2549009 promoter variants haplotyping was performed on a total of 41 healthy Caucasians. Two independent read out assays for ASTQ analysis, (i) differently labelled allele-specific 5' nuclease assays and (ii) RFLP analyses, were conducted on material from IRF-1 rs839 heterozygous individuals.

Genotype distribution of IRF-1 rs839 was found to be in line with HWE, and MAF (0.366) was found to be close to that given at NCBI. On the basis of these findings, LD was calculated using the software HaploView version 4.1. With $D' = 0.947$ and $r^2 = 0.897$ the IRF-1 exon 10 SNP rs839 was found to be a suitable polymorphism for assigning IRF-1 transcripts to respective IRF-1 rs2549009 promoter variants. The wild-type alleles G and variant alleles A for both SNPs, IRF-1 rs2549009 and IRF-1 rs839, were found to be in a high LD (Mertens et al., 2010).

Some of the 16 IRF-1 rs839 heterozygous individuals showed an allelic expression imbalance for IRF-1 mRNAs. (i) The rs839 A/G ratios varied from 0.84 ± 0.04 (donor #12) to 1.13 ± 0.13 (donor #9) and similar results were detected by (ii) RFLP, where IRF-1 rs839 A/G ratios varied between 0.87 and 1.06. Linear regression analysis of data derived from the two different read out assays revealed a close correlation with $p = 0.003$ and $r = 0.71$. These results provide first evidence for a bidirectional IRF-1 gene expression imbalance in PBMC (ASTQ experiments were performed by our research team members; Mertens et al., 2010).

Further, the relative proportion of IRF-1 rs2549009 promoter variants bound to the transcriptionally active ser2-P RNAPII was found to correlate significantly ($p = 0.0022$; $r = 0.99$) to respective allele-specific IRF-1 transcripts (Fig. 23). In detail, donor #9 showed an IRF-1 rs2549009 A/G ratio of 1.22 in H5-immunoprecipitated material, indicating a preferential binding of the transcriptionally active RNAPII to the IRF-1 rs2549009 A promoter variant. The same donor also exhibited an IRF-1 rs839 A/G transcript ratio of 1.13 in PBMC. Donors #1 and #4 showed higher amounts of both IRF-1 rs2549009 G promoter variants (IRF-1 rs2549009 A/G ratio 0.80; 0.86) and respective IRF-1 rs839 G transcripts (IRF-1 rs839 A/G ratio 0.92; 0.90). Almost no difference in the amount of allelic IRF-1 promoter variants and respective transcripts were found in PCMC from donors #8 and #10 (Mertens et al., 2010).

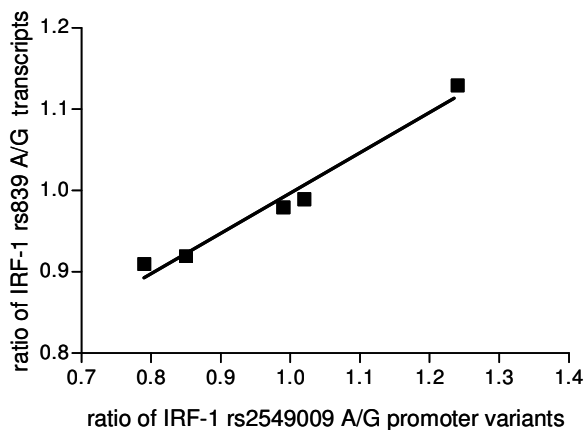


Figure 23: Close correlation between IRF-1 rs2549009 promoter variants and respective IRF-1 rs839 transcripts

The relative binding of the active ser2-P RNAPII to IRF-1 rs2549009 promoter variants was found to correlate significantly to respective allelic IRF-1 transcripts in PBMC from 5 healthy double-heterozygous (IRF-1 rs2549009, IRF-1 rs839) Caucasians (Modified from Mertens et al., 2010).

The relative binding frequency of the transcriptionally active RNAPII to IRF-1 promoter variants (i) does not seem to be solely dependent on IRF-1 rs2549009 variants and (ii) was found to correlate to respective IRF-1 transcripts in freshly isolated PBMC from healthy Caucasians. These provide evidence for an IRF-1 gene expression imbalance that does not appear to be solely controlled by the IRF-1 rs2549009 promoter variant in PBMC.

Allelic gene expression has been shown for up to 5% of human autosomal genes (Pastinen et al., 2005) and whole genome linkage studies of total gene expression have suggested that 1-3% of human genes harbour common *cis*-acting variants (Buckland, 2004; Rockman and Wray, 2002; Serre et al., 2008). The small bidirectional imbalance in allele-specific IRF-1 promoter activity and gene transcription, assessed by HaploChIP and ASTQ, was found to be dependent on the respective donors. This finding is quite in accordance with observations by others. It has been reported that imbalances of a particular gene may only occur in some individuals (Buckland, 2004; Rockman and Wray, 2002). Further, also the ranges of allelic imbalances were found to vary a lot, not only between genes but also between individuals. Differences of less than 10% to 4.3-fold or higher have been reported (Buckland, 2004; Serre et al., 2008; Southam et al., 2007; Yan et al., 2002).

3.2.3. Further functional studies on IRF-1 rs2549009 gene expression

A parallel analysis of total IRF-1 gene expression revealed that total IRF-1 gene expression was not significantly different when related to IRF-1 rs2549009 promoter genotypes ($p = 0.2709$) in 41 healthy Caucasians. A more detailed analysis within the group of 16 individuals heterozygous in IRF-1 rs2549009, however, revealed a close and positive relation ($p = 0.0035$, $r = -0.68$) between the relative amount of G allele transcripts and total IRF-1 mRNA levels in PBMC (these analyses were done in cooperation with our research team members; Mertens et al., 2010).

3.2.4. LD between IRF-1 rs2549009 and neighbouring promoter polymorphisms

The bidirectional allele expression imbalance in IRF-1 rs2549009 heterozygotes (Mertens et al., 2010) indicates a gene expression regulation that is not solely controlled by IRF-1 rs2549009. IRF-1 gene expression might be regulated by other genetic variants in *cis* or might be influenced by yet unidentified gene-environment interactions in *trans* (Mertens et al., 2010).

Two other SNPs, IRF-1 rs2549007 and IRF-1 rs2706384 located 110 nt and 115 nt upstream of IRF-1 rs2549009 within the IRF-1 promoter region, have been described by Saito et al. as being in a high LD with IRF-1 rs2549009 in a Japanese population (Saito et al., 2002).

To address the question of whether these two SNPs might have any *cis*-regulatory relevance for IRF-1 gene expression in Caucasians, geno-/haplotyping was performed on gDNA samples from 40 individuals of the healthy cohort.

Genotype distribution of IRF-1 rs2549007 among the large healthy cohort was found to be in line with HWE, also MAF of IRF-1 rs2549007 was found to be similar to IRF-1 rs2549009 (see 3.2. and Tab. 4). On this basis LD was calculated again using the software HaploView version 4.1 and with $D' = 1.0$ and $r^2 = 0.947$, these two SNPs were evaluated as being in a high LD. Therefore, any observations that were made for IRF-1 rs2549009 might be as well attributed to IRF-1 rs2549007 (Mertens et al., 2010).

Genotype distribution of IRF-1 rs2706384 was also found to be in line with HWE and MAF of IRF-1 rs2706384 was found to be similar to IRF-1 rs2549009 (see 3.2. and Tab. 4) among the healthy cohort. Allelic LD for IRF-1 rs2549009 and IRF-1 rs2706384 was also calculated on this basis using the software HaploView version 4.1 and found to be close ($D' = 0.926$; $r^2 = 0.604$). Thus, observations made for IRF-1 rs2549009 seem to be attributable to IRF-1 rs2706384 as well (Mertens et al., 2010).

Tab. 4: Genotype distributions among the healthy Caucasian cohort

	ww	wv	vv	MAF	p^a
rs2549007 [G/A]	18 (45.0%)	15 (37.5%)	7 (17.5%)	0.362	0.302
rs2706384 [A/C]	25 (62.5%)	8 (20.0%)	7 (17.5%)	0.275	0.154

^a p exact test for Hardy-Weinberg equilibrium

w = wild-type; v = variant

As expected, the two IRF-1 promoter SNPs rs2549007 and rs2706384 were found to be in a close LD to IRF-1 rs2549009 in our cohort, as well. Thus, at least an impact of these two other SNPs on IRF-1 gene expression can be ruled out.

A contribution of gene-environment interaction to IRF-1 rs2549009-dependent gene expression is supported by comparing IRF-1 gene expression with its inducibility by IFN- α in *ex vivo*-derived PBMC from healthy and HCV-infected individuals (Mertens et al., 2010).

As mentioned before, IRF-1 mRNA expression levels were not related to IRF-1 rs2549009 promoter genotypes in healthy individuals (see 3.2.3.).

Also gene expression levels in *ex vivo*-derived PBMC from patients with chronic HCV infection were not found to be significantly different when compared to IRF-1 rs2549009 genotypes. But total IRF-1 mRNA levels in PBMC from chronic hepatitis C patients were significantly higher than in healthy individuals (Aidery, 2010). However, responsiveness of IRF-1 gene expression in PBMC to human recombinant IFN- α *in vitro* was found to be significantly higher in healthy individuals and, interestingly, more pronounced in IRF-1 rs2549009 AA homozygotes (Aidery, 2010).

Thus, the findings of a positive relation of allelic and total IRF-1 gene expression (Aidery, 2010; Mertens et al., 2010) suggests a biological impact of IRF-1 rs2549009 on IRF-1 gene expression, that seems also to be controlled by a positive or negative regulation in *trans* (Mertens et al., 2010).

4. Summary

The IRF-1 promoter SNP rs2549009 and the variant rs2569190 within the CD14 gene have been described as being associated with various disease conditions. *In vitro* experiments (EMSA; reporter gene assays) showed that the T allele of CD14 rs2569190 favours transcription factor binding and that DNA elements containing the respective variant allele promote gene transcription. The IRF-1 rs2549009-carrying DNA element induces allele-specific constitutive expression levels of a luciferase reporter in various cell lines. To determine whether these SNPs confer any functional effect on transcriptional activity in their natural genomic environment, analyses of allele-specific promoter activities were conducted by applying HaploChIP on *ex vivo*-derived PBMC. Therefore, antibodies were used that are directed against ser5-P and ser2-P isoforms of RNAPII that reflect binding and enzymatic activity at initiation and elongation stages of the transcription cycle, respectively.

Ser5-P RNAPII was found to be preferentially recruited to the variant alleles of CD14 rs2569190 as well as of IRF-1 rs2549009, results that are in line with those from the above-mentioned *in vitro* experiments.

The active RNAPII, on the contrary, was found to not be preferentially bound to the wild-type or variant allele of CD14 rs2569190. This finding is supported by (i) the absence of a CD14 mRNA expression imbalance and (ii) the absence of any relation of three other SNPs within the CD14 gene and allelic gene expression. Sequencing of the CD14 gene also failed to reveal any so-far unknown SNP that could influence CD14 promoter activity. The genetic variant CD14 rs2569190 thus might be advantageous for transcription initiation. Endotoxin sensitivity associated with this polymorphism, however, does not appear to rely on allele-specific CD14 gene transcription at least in PBMC (Mertens et al., 2009).

Regarding IRF-1 rs2549009, a preferential binding of the ser2-P RNAPII to the G or A promoter variants or binding without any preference - depending on the donor - was observed in PBMC from healthy Caucasians. This finding actually accords with the varying results from different reporter gene assays and may explain the variance. Furthermore, the relative binding frequency of the active RNAPII to IRF-1 rs2549009 promoter variants was found to be closely related to the relative amount of allele-specific IRF-1 transcripts. These results provide evidence for a bidirectional IRF-1 gene expression that does not appear to be solely controlled by IRF-1 rs2549009 in *cis*, but may rely on a yet undetected polymorphism or haplotype or on environmental control in *trans* (Mertens et al., 2010).

5. Supplementary material

5.1. Genotyping

CD14	
rs2569190 (C/T)	
forward primer	5' CTA GAT GCC CTG CAG AAT CCT 3'
reverse primer	5' CCC TTC CTT TCC TGG AAA TAT TGC A 3'
probe T	VIC 5' CCT GTT TAC GG <u>T</u> CCC CCT 3'
probe C	FAM 5' CT GTT TAC GG <u>C</u> CCC CCT 3'
rs5744455 (C/T)	
rs574 tetra T f	5' AAG GAA GGG GGA ATT TTT CTT TAG CC <u>T</u> 3'
rs574 tetra C r	5' GGT AGA ATT AGG TTC AAG AAA AGG AAG GT <u>G</u> 3'
rs574 tetra for	5' CTG AGC AAC AGA GCA AGA CTC TAT CTC A 3'
rs574 tetra rev	5' CAT TTC TTT GAC TTC TTC CTT GTC TTG G 3'
rs2563298(G/T)	
rs2563 tetra T r	5' TAA AGG TCT GTT AAA TGA ATG ACA CGA AA <u>A</u> 3'
rs2563 tetra G f	5' CCC ACC TTT ATT AAA ATC TTA AAG AAC AG <u>G</u> 3'
rs2563 tetra for	5' TTT GCC TAA GAT CCA AGA CAG AAT AAT G 3'
rs2563 tetra rev	5' TTT CTT AGG GAG TTA GGA TGA AGA AAG C 3'
IRF-1	
rs2549009 (G/A)	
forward primer	5' GGA CAA GGC GGA GTG AGA 3'
reverse primer	5' GGC TGC CCG GGA CAG 3'
probe A	FAM 5' CCC C <u>T</u> G GCC CCA G 3'
probe G	VIC 5' CCC C <u>G</u> G GCC CCA G 3'
rs2549007 (G/A)	
forward primer	5' AGG CTT TCT GCC TTC TTC ACT T 3'
reverse A primer	5' GGT ATA TCT CCC GAA CGC ATG <u>T</u> 3'
reverse G primer	5' GTA TAT CTC CCG AAC GCA TG <u>C</u> 3'

rs2706384 (A/C)	
reverse primer	5' CTG TCC TCT CAC TCC GCC TTG TCC 3'
forward C primer	5' CAA GTG CCC GGG CGA CC _C 3'
forward A primer	5' CAA GTG CCC GGG CGA CC _A 3'
SNRPN	
rs705	
Detailed information not available. TaqMan [®] Genotyping Assays C_2066555_10	

5.2. HaploChIP

CD14	
primer for RFLP analysis	
rs2569190 for	5' GACACAGAACCCTAGATGCCCT 3'
rs2569190 rev	5' GTGAACTCTTCGGCTGCCTC 3'
enzymes for RFLP analysis	
<i>TaaI</i> allele T	5' ACN↓GT 3' 3' TG↑NCA 5'
<i>HaeIII</i> allele C	5' GG↓CC 3' 3' CC↑GG 5'
IRF-1	
primer for RFLP analysis	
IRF rs254 for	5' CAG GAG GGT GAA AAG ATG GCC 3'
IRF rs254 rev	5' CGA GCG CTC CCA ATC CAC 3'
enzymes for RFLP analysis	
<i>MvaI</i> allele A	5' CC↓AGG 3' 3' GG↑TCC 5'
<i>BcnI</i> allele G	5' CC↓CGG 3' 3' GG↑CC 5'

5.3. Sequencing

CD14	
primer for region 1	
for1	5' GCC GAG ATC ATG GCA 3'
rev1	5' ACC CTG ATC ACC TCC CCA C 3'
primer for region 2	
for2	5' CCC CTC CCT GAA ACA TCC TT 3'
rev2	5' GAG TGT GCT TGG GCA ATG CT 3'
primer for region 3	
for3	5' GGA CTT GCA CTT TCC AGC TTG 3'
rev3	5' GCA CAT AGC AGA CAT CCA ATA AAG G 3'

```

GAGGCAGAGAATTGCTTGAACCCAGGAGGCGGAGGTTGCACTGAGCCGAGATCATGGCAC
TCTACTCCAGCCTGAGCAACAGAGCAAGACTCTATCTCAAAAAAAAAAAAAAAAAAGAAA
GAAAGTTATTACTTAATCAAAGGAGCAAGGAAAAAAAAAGGAAGGGGAATTTTCTTTA
GACTAACTTCTTTCTTGAACCTAATCTACCCCTTGGTGCCAAACAGATGAGGTTCA
rs5744455 (C/T)
CAATCTCTTCCACAAAACATGCAGTTAAATATCTGAGGATATTGAGGACTTGGATTTGG
TGGCAGGAGATCAACATAAACCAAGACAAGGAAGAAGTCAAAGAAATGAATCAAGTAGAT
TCTCTGGGATATAAGGTGGGGGATTGGGGGTTGATAGTGCAGAGTATGGTACTGGCC
TAAGGCACTGAGGATCATCTTTTCCACACCCACAGAGAGGCTTAGGCTCCCGAGTC
AACAGGGCATTACCCCTGGGGCGCCTGAGTCATCAGGACACTGCCAGGAGACACAGAA
CCCTAGATGCCCTGCAGAACTCTTCTGTTACGGCCCCCTCCCTGAAACATCCTTCATT
rs2569190 (C/T)
GCAATATTTCCAGGAAGGAAGGGGGCTGGCTCGGAGGAAGAGAGGTGGGGAGGTGATCA
GGGTTCACAGAGGAGGGAAGTGAATGACATCCAGGATTTCATAAACTGTCAGAGGCAGC
CGAAGAGTTTCAAGTGTGAAGCCTGGAAGCCGGCGGGTCCCGCTGTGTAGGAAAGAAAS
TAAAGCACTTCCAGAGCCTGTCCGGAGCTCAGAGGTTCCGAAGACTTATCGACCATG
AGTGTAGGGTCTTGGGGTCAACGCGTGCCTCTCGGAGCCACAGGGGTTGGATGGGGCC
TCCTAGACCTCTGCTCTCTCCCCAGGAGCGCGCTCTGTGCTGTTGCTGCTGCTGCC
GCTGGTGACGCTCTCTGCGACCCAGCCAGAACCTTGTGAGCTGGACGATGAAGATTTCG
CTGCGTCTGCAACTTCTCCGAACCTCAGCCGACTGGTCCGAAGCCTTCCAGTGTGTGTC
TGCGTAGAGGTGGAGATCCATGCCGGCGGTCTCAACCTAGAGCCGTTTCTAAAGCGCGT
CGATGCGGACGCGCACCCGCGCAGTATGCTGACAGGTCAGGCTCTCCGCGTGCAGCG
GCTCAGAGTGGGAGCCGACAGGTTCTGCTCAGCTACTGTTAGGCGCCCTGCGTGTGCT
AGCGTACTCCCGCTCAAGGAAGTACGCTCGAGGACCTAAAGATAACCGGCACCATGCC
TCCGCTGCCCTCTGGAAGCCACAGGACTTGCACTTTCCAGCTTGCGCCTACGCAACGTGTC
GTGGGCGCAGGGCGTTCTTGGCTCGCCGAGCTGCAGCAGTGGCTCAAGCCAGGCTCA
GGTACTGAGCATTGCCAAGCAGCTCGCTGCTTTTCTGCGAACAGGTTGCGCGCTT
CCCGGCCCTTACCAGCTAGACCTGTCTGACATCCTGGACTGGGCGAACCGGAGTAT
GGCGGCTCTCTGTCCCAAGTTCCCGGCCATCCAGAACTAGCGCTGCGCAACACAGG
AATGGAGACGCCACAGGCGTGTGCGCGCAGTGGCGGCGCAGGTGTGACGCCCCACAG
CCTAGACCTCAGCCACAACTCGTGCAGCCACCGTAAACCTAGCGCTCCGAGATGCAT
GTGTCCAGCGCCCTGAAGTCCCTCAATCTGTCTGCTGCTGGCTGGAACAGGTGCCTAA
AGGACTGCCAGCCAAGCTCAGAGTGTCTGATCTCAGTGCAGAGTGAACAGGGCGCC
GCAGCCTGACGAGCTGCCGAGGTGGATAAAGTACACTGGACGGGAATCCCTTCTGGT
CCCTGGAAGTGCCTTCCCCACAGGGCTCAATGAAGTCCGGCGTGGTCCAGCCTGTGC
ACGTTCCAGCCTGTGCGTGGGGGTGTGCGGAACCTGGTGTCTTCCAGGGGCCCCGGGG
CTTTGCCTAAGATCCAGACAGAATAATGAATGAGTCAAACTGCCCTTGGCTTCCAGGGGA
GTCCCGTCAAGACGTTGAGGACTTTTCGACCAATTCAACCCCTTGGCCCCACCTTTATTAA
AATCTTAAASAACGCTCCGTTGCATTTAATTAACAGACCTTTATTGGATGTCTGCTATG
rs4914 (C/G)
TGC TGGGCACAGTGTGGATGGGGAATTCATCATACTAGGCTTTCTTCATCCTAACTCC

```

Figure 24: Sequenced region of the CD14 gene and SNP positions within the gene and transcript

CD14 sequence and SNP information are taken from ensemble (ENSG00000170458) and NCBI. Transcribed region (transcript variant 1: NM_000591.2) is indicated in orange. Flanking sequencing primers are given in grey and overlapping regions are marked in yellow. Known polymorphisms are given in green, while SNPs with a MAF > 0.1 are additionally marked in blue. During sequencing of 10 different CD14 rs2569190 heterozygous individuals possible unknown variant positions within the promoter region were detected and are indicated in red.

5.4. Haplotyping

CD14	
Allele-specific primer rs5744455 (C/T)	
rs574 tetra T f	5' AAG GAA GGG GGA ATT TTT CTT TAG CCT <u>T</u> 3'
rs574 tetra C f	5' AAG GAA GGG GGA ATT TTT CTT TAG CC <u>C</u> 3'
CD14 tetra 3-rev	5' TTT TCT TGA GGA GGA CAG ATA GGG TTT C 3'
Allele-specific primer rs2569190 (C/T)	
rs2569190 T f	5' AAG GAA GGG GGA ATT TTT CTT TAG CCT <u>T</u> 3'
rs2569190 C f	5' AAG GAA GGG GGA ATT TTT CTT TAG CC <u>C</u> 3'

6. Publications

6.1. Functional impact of endotoxin receptor CD14 polymorphisms on transcriptional activity

Jasmin Mertens¹, Rusudan Bregadze¹, Ashham Mansur¹, Eva Askar¹, Heike Bickeböllner², Giuliano Ramadori¹, Sabine Mihm¹

¹Division of Gastroenterology and Endocrinology, Department of Internal Medicine, ²Department of Genetic Epidemiology,

Georg-August-Universität, D-37075 Göttingen, Germany

Published in: Journal of Molecular Medicine (2009) 87: 815-824

DOI 10.1007/s00109-009-0479-7

Author contribution:

This study was designed and supervised by SM and GR. JM established and performed HaploChIP assays and carried out sequence analyses. RB and JM established and performed geno-/haplotyping. RB carried out the ASTQ analyses. JM and RB were involved in writing the manuscript, i.e. JM designed the figures and RB prepared the tables. HB carried out the statistical analyses.

J Mol Med (2009) 87:815–824
DOI 10.1007/s00109-009-0479-7

ORIGINAL ARTICLE

Functional impact of endotoxin receptor CD14 polymorphisms on transcriptional activity

Jasmin Mertens · Rusudan Bregadze ·
Ashham Mansur · Eva Askar · Heike Bickeböllner ·
Giuliano Ramadori · Sabine Mihm

Received: 23 February 2009 / Revised: 17 April 2009 / Accepted: 23 April 2009 / Published online: 27 May 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract The polymorphism rs2569190 within the CD14 endotoxin (lipopolysaccharide, LPS) receptor gene is associated with various disease conditions that are assumed to rely on endotoxin sensitivity. In vitro experiments suggest that the T allele sensitizes the host for exogenous or endogenous LPS via an enhanced CD14 expression. To prove the impact of this single nucleotide polymorphism in its natural genomic context in vivo, two parameters of gene transcription were analyzed in peripheral blood mononuclear cells (PBMC) from single healthy individuals: (a) recruitment of RNA polymerase II by haplotype-specific chromatin immunoprecipitation and (b) the relative amount of transcripts by allele-specific transcript quantification (ASTQ). RNA polymerase II was found to be twice as much bound to the most prevalent haplotype, C-T-C-G, the only one carrying a T at the position rs2569190 of interest. ASTQ employing two independent read-out assays revealed, however, similar transcript numbers originating from C-T-C-G and non-C-T-C-G haplotypes. Total CD14

mRNA levels from freshly isolated PBMC, moreover, were neither related to donors' geno- nor haplogenotypes. Our data argue for a functional impact of the rs2569190 polymorphism in terms of a stronger transcription initiation on T allele gene variants even if preferential allele-specific binding does not result in an increase in transcript numbers. Endotoxin sensitivity associated with this genetic variation appears not to rely solely on a cis-acting regulatory impact of rs2569190 on CD14 gene transcription in PBMC.

Keywords SNP · LPS · Gene polymorphism · Gene expression · Innate immunity

Introduction

A host might get into contact with bacterially derived endotoxin (lipopolysaccharide, LPS) at the respiratory endothelium via environmental exposure of LPS containing dust, at the intestinal endothelium in the gut via the commensal microflora, at liver sinusoids via the portal vein, or in extravascular compartments when infection has passed the microvascular endothelium. In mammals, endotoxin sensing is mediated by circulating LPS-binding protein (LBP) which transfers LPS to a signaling receptor complex composed of a glycosylphosphatidylinositol-anchored membrane-associated form of CD14 (mCD14), MD-2, and Toll-like receptor-4 (TLR4) on myeloid cells or by LBP and a soluble form of CD14 (sCD14) which transfers LPS to a MD-2/TLR4 receptor complex on nonmyeloid cells. The soluble form of CD14 is either shedded from cells bearing mCD14 or from direct secretion. Since sCD14 is also able to compete with mCD14 for LPS, a dual role for sCD14 has been proposed, namely an inhibitory function on systemic responses by

J. M. and R. B. contributed equally to this work.

J. Mertens · R. Bregadze · A. Mansur · E. Askar · G. Ramadori ·
S. Mihm
Department of Gastroenterology and Endocrinology,
University Medical Center Göttingen,
37075 Göttingen, Germany

H. Bickeböllner
Department of Genetic Epidemiology, Georg-August-Universität,
37075 Göttingen, Germany

S. Mihm (✉)
Department of Gastroenterology and Endocrinology,
University Medical Center Göttingen,
Robert-Koch-Straße 40,
37075 Göttingen, Germany
e-mail: smihm@med.uni-goettingen.de

promoting LPS efflux from mCD14 and transferring it to plasma lipoproteins and a stimulatory function on local inflammation in tissues that may increase resistance to bacteria [1]. Apart from that, LBP, CD14, and MD-2 have been understood to function as upstream accessory molecules that mediate physical ligand presentation, i.e., the extraction of endotoxin monomers from the bacterial outer membrane, and determine the amplitude of TLR4 activation (reviewed in [2]).

Aside from genetic variations within the TLR4 gene which have been shown to be associated with individual endotoxin sensitivity [3, 4], many lines of evidence suggest that also genetic variations within the CD14 gene do determine the outcome of LPS exposure. For instance, healthy individuals with the TT genotype at the variant position rs2569190 have higher serum baseline levels of sCD14 than those with the CC genotype [5], TT homozygous children who were skin test positive to local aeroallergens have significantly lower total IgE levels when compared to C allele carriers [6], and in Finish children, a gene–environment interaction between T allele carriers and *Helicobacter pylori* antibodies on serum total IgE was shown [7]. The same polymorphism has been shown to be associated with alcoholic or cryptogenic chronic liver disease progression [8–11]; it is supposed that the T allele confers susceptibility to endogenous gut-derived LPS via an enhanced expression of CD14 on liver cells thus promoting progression of a chronic liver disease. By applying electrophoretic mobility shift assays (EMSA), this SNP has been found to lie within a GC box that contains a binding site for Sp1, Sp2, and Sp3 transcription factors. A DNA element containing the T allele was found to favor transcription factor binding. Moreover, luciferase reporter gene assays revealed the T allele DNA element to promote gene transcription in monocytic cells via an enhanced affinity to Sp1 and Sp2 [6, 12].

This study aimed at investigating of whether the variant position rs2569190 confers allele-specific gene expression also in a natural genomic context, i.e., whether it can be regarded as a regulatory polymorphism. The majority of SNPs which influence illness are assumed to be regulatory (rSNP) meaning that they lie outside the amino acid coding regions of genes and modify the expression level of transcripts [13]. By taking further SNPs within the CD14 gene and nearby regions with relevant minor allele frequencies (MAF; >0.1) into account, two parameters of gene transcription were analyzed in freshly prepared human peripheral blood mononuclear cells (PBMC). First, haplotype-specific chromatin immunoprecipitation (Haplo-ChIP) using antibodies directed against serine 5 phosphorylated RNA polymerase II was applied to assess transcription initiation [14, 15]. Second, allele-specific transcript quantification (ASTQ) was applied to quantify

the relative numbers of transcripts that have been generated from the two allelic gene variants. Both techniques circumvent intrinsic errors in interindividual comparisons, since comparison of alleles is made within rather than between samples [16]. Nonetheless, total numbers of CD14 transcripts were also assessed with regard to the blood donors' geno- and haplotypes.

Materials and methods

Blood donors

Samples of 30 ml venous blood were taken from 42 healthy Caucasian blood donors (20 female, 22 male; mean age 31.8 years, range 23–57) who gave a written informed consent. The study was approved by the local ethical committee and conformed to the ethical principles of the 2000 Declaration of Helsinki.

HaploChIP [17]

Chromatin immunoprecipitation was conducted using the ChIP-IT Express kit (Active Motif, Rixensart, Belgium) according to the supplier's protocol. In brief, $1\text{--}10\times 10^6$ cells were cross-linked using 0.37% formaldehyde for 6 min at room temperature. Fixation reactions were stopped by adding glycine. Cells were washed once in ice-cold PBS, resuspended in ice-cold lysis buffer, and subjected to dounce homogenization to release the nuclei. Pelleted nuclei were resuspended in digestion buffer and incubated in an enzymatic shearing cocktail for 10 min at 37°C. After centrifugation, sheared chromatin was stored in aliquots at -80°C until further use for immunoprecipitation or as control "input" DNA.

Chromatin corresponding to 1×10^6 cells was incubated with an antibody directed against serine 5 phosphorylated RNA polymerase II or with a control mock (IgG) antibody, and antibody-bound protein/DNA complexes were precipitated through use of magnetic protein G-coated beads. The captured chromatin was then subjected to reverse cross-linking and protease K digestion and taken up in a final volume of 100 μl .

Input DNA (0.8 μl) and antibody captured DNA (8 μl) were amplified by polymerase chain reaction (PCR) in 100 μl reactions using the primers rs2569190 for 5'-GAC ACA GAA CCC TAG ATG CCC T-3' and rs2569190 rev 5'-GTG AAC TCT TCG GCT GCC TC-3'. Cycling conditions were a 2-min denaturation at 94°C , 28 cycles of 30 s denaturation at 94°C , and 1 min annealing and elongation at 60°C followed by a final elongation step for 7 min at 60°C . PCR products were subjected to restriction fragment length polymorphism (RFLP) analysis. The

products (201 bp) were digested with *HaeIII* (Fermentas, St. Leon-Rot, Germany) overnight. Samples were analyzed in the Agilent 2100 bioanalyzer using the DNA 1000 LabChip kit (Agilent Technologies, Böblingen, Germany) according to the supplier's instructions.

Preparation of PBMC and isolation of total cellular RNA

PBMC from approximately 30 ml of heparinized peripheral blood were prepared by Ficoll density centrifugation using guanidinium isothiocyanate, and total cellular RNA was isolated by CsCl density gradient ultracentrifugation essentially as described [18].

ASTQ

Discrimination of transcripts using rs4914 as a marker SNP was achieved by two different read-out assays. First, discrimination and quantification of transcript variants were achieved by real-time PCR using two differently fluorescence dye-labeled allele-specific minor groove binder (MGB) probes for the SNP rs4914 (VIC 5'-CCT TGG AGG AGC AC-3', FAM 5'-CTT GGA GCA GCA CC-3') and gene-specific primers (rs4914 for 5'-CTG AAG CCA AGG CAG TTT GAG T-3', rs4914 rev 5'-TCC CAG CCT GTG CAC GTT-3'). The assay was validated to be allele-specific (absence of a signal from the second probe in homozygote genomic DNA (gDNA) samples) and quantitative as judged from reaction efficiencies and from comparable signals in heterozygote gDNA samples (data not shown).

Secondly, discrimination and quantification of transcript variants was carried out by RFLP analyses: A product of 201 bp in size was generated by conventional PCR (CD14 exon2 for 5'-GCC CGA GGT GGA TAA CCT G-3', CD14 exon2 rev 5'-GTC CAT TCA TTA TTC TGT CTT GGA TC-3'), purified and concentrated using the PrepEase gel extraction kit (USB, Cleveland, OH, USA), and subjected to digestion by *SatI* and *Alw21I* (Fermentas) for 2 h. *SatI* cleaves C allele products into fragments of 55 and 146 bp in size, whereas digestion by *Alw21I* leads to fragments of 106 and 95 bp in size for C allele products and to fragments of 106, 50, and 40 bp in size for G allele amplicons, respectively. Samples were separated and analyzed in the Agilent 2100 bioanalyzer using the DNA 1000 LabChip kit (Agilent Technologies). The relative amount of C and G allele products and of hybrids was calculated.

Sequencing

Three overlapping regions of the CD14 gene and nearby regions spanning 2,101 nucleotides (Fig. 1) were amplified using gDNA as template, GoTaq Colorless Master Mix

(Promega GmbH, Mannheim, Germany), and 0.36 μ M of primers each (for1 5'-GCC GAG ATC ATG GCA-3', rev1 5'-ACC CTG ATC ACC TCC CCA C-3'; for2 5'-CCC CTC CCT GAA ACA TCC TT-3', rev2 5'-GAG TGT GCT TGG GCA ATG CT-3'; for3 5'-GGA CTT GCA CTT TCC AGC TTG-3', rev3 5'-GCA CAT AGC AGA CAT CCA ATA AAG G-3'). Products were purified following the spin protocol of the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) and sequenced using an automated sequencer (SeqLab, Göttingen, Germany). Alignments and analyses were done using the Basic Local Alignment Search Tool program of National Center for Biotechnology Information (NCBI; National Institutes of Health, Bethesda, MD, USA) and version 2.01 of the software Chromas LITE (Technelysium Pty Ltd, Tewantin, QLD, Australia).

Genotyping

Genomic DNA was purified from whole blood samples using the QIAamp DNA Blood Mini Kit following the blood and body fluid spin protocol (Qiagen). Typically, genotype determinations were performed by using 10 ng of PBMC-derived gDNA.

Genotyping of the variant position of interest, rs2569190, was carried out by allelic discrimination in 5'-nuclease real-time PCR assays in the sequence detection system ABI prism 7000 (Applied Biosystems, Darmstadt, Germany) according to the supplier's instructions. In brief, gDNA was amplified with 0.36 μ M of primers each (forward 5'-CTA GAT GCC CTG CAG AAT CCT T-3', reverse 5'-CCC TTC CTT TCC TGG AAA TAT TGC A-3') using the TaqMan Universal Master Mix (Applied Biosystems). Allelic discrimination was achieved by adding 0.2 μ M differentially fluorescence dye-labeled allele-specific MGB probes (rs2569190: VIC 5'-CCT GTT ACG GTC CCC CT-3', FAM 5'-CTG TTA CGG CCC CCC T-3').

Genotyping of the variant position rs4914, the marker SNP for ASTQ, was performed in the sequence detection system ABI prism 7000 by using a commercially available TaqMan genotyping assay (C_8724832_10) according to the supplier's instructions. Discrimination of genomic DNA and transcript variants at SNP rs705 within the imprinted gene small nuclear ribonucleoprotein polypeptide N (SNRPN) was carried out using a commercially available genotyping assay from ABI (C_2066555_10).

Genotyping of the SNPs rs5744455 and rs2563298 was performed by tetra-primer amplification refractory mutation system (ARMS) PCR as introduced by Ye et al. [19]. In brief, gDNA was used as template in conventional amplification reactions using 0.18 μ M of each primer (primer rs574 tetra T for 5' AAG GAA GGG GGA ATT TTT CTT TAG CCT 3', primer rs574 tetra C rev 5' GGT AGA ATT AGG TTC AAG AAA AGG AAG GTG 3',

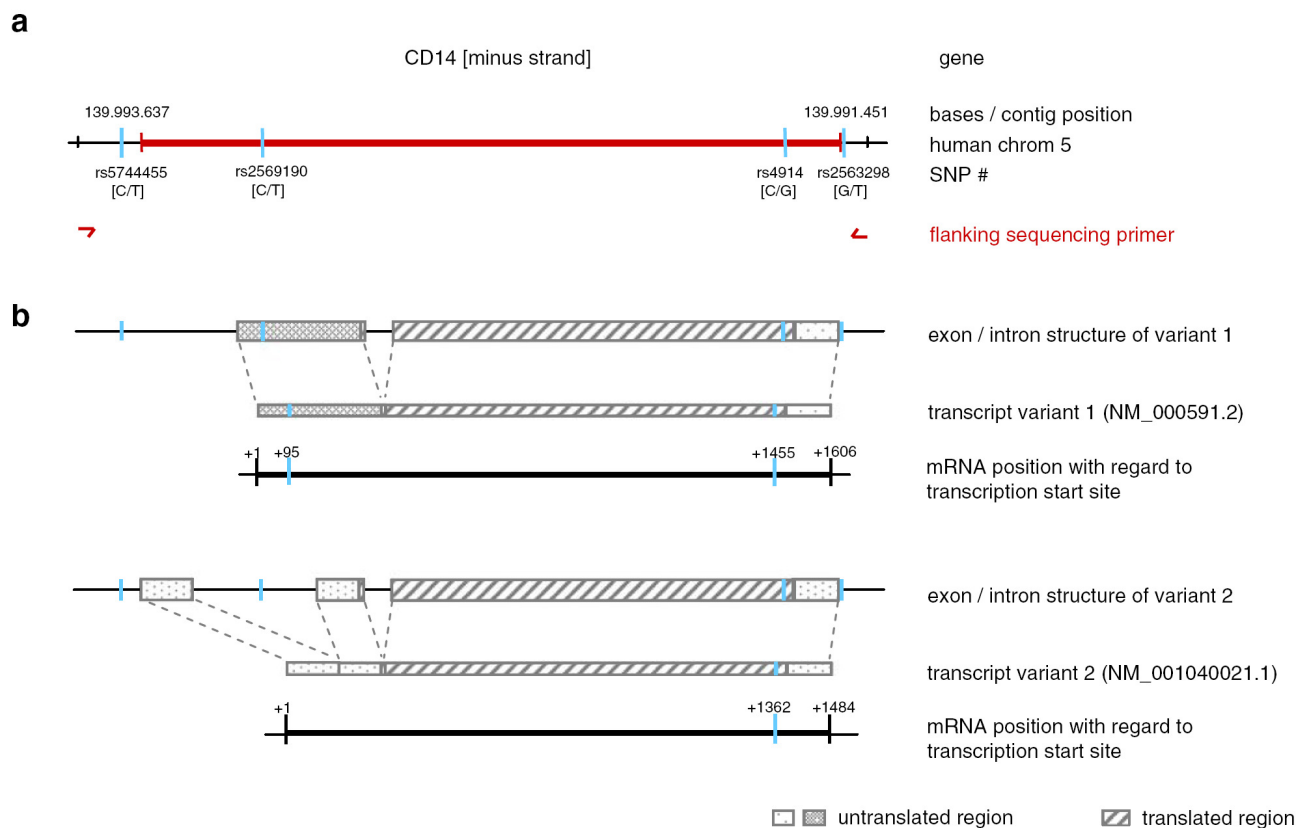


Fig. 1 Schematic presentation of the structure of the CD14 gene and of the two transcript variants. **a** In humans, the CD14 gene is located on chromosome 5. The gene and the nearby 5' and 3' regions contain four SNPs with MAFs >0.1 as indicated. Presence of these four SNPs was confirmed by sequencing the indicated region. **b** Differential

splicing results in the existence of two transcript variants. Both of them do contain rs4914, the marker SNP used for ASTQ. In transcript variant 1, rs2569190 is located within the 5'UTR; because of its intronic location in variant 2, rs2569190 is absent from transcript variant 2

primer rs574 tetra for 5' CTG AGC AAC AGA GCA AGA CTC TAT CTC A 3', primer rs574 tetra rev 5' CAT TTC TTT GAC TTC TTC CTT GTC TTG G 3' for rs5744455, and primer rs2563 tetra G for 5' CCC ACC TTT ATT AAA ATC TTA AAG AAC AGG 3', primer rs2563 tetra T rev 5' TAA AGG TCT GTT AAA TGA ATG ACA CGA AA 3', primer rs2563 tetra for 5' TTT GCC TAA GAT CCA AGA CAG AAT AAT G 3', and primer rs2563 tetra rev 5' TTT CTT AGG GAG TTA GGA TGA AGA AAG C 3' for rs2563298). Amplicons were separated in 4% agarose gels. Product patterns of 277/142, 277/142/192, and 277/192 bp correspond to CC, CT, and TT genotypes of rs5744455 and patterns of 247/143, 247/143/163, and 247/163 bp to GG, GT, and TT rs2563298 genotypes, respectively.

Haplotyping

Haplotyping of the four respective SNPs was performed by a combination of ARMS-PCR and genotyping assays. Amplicons spanning the four SNPs of interest were generated using rs5744455 and rs2569190 allele-specific

primers (rs574 tetra T for, rs574 tetra C for 5'-AAG GAA GGG GGA ATT TTT CTT TAG CCC-3', CD14 tetra3-rev 5'-TTT TCT TGA GGA GGA CAG ATA GGG TTT C-3', and rs2569190 C for 5'-CCT GCA GAA TCC TTC CTG TTA CTG C-3', rs2569190 T for 5'-CCC TGC AGA ATC CTT CCT GTT ACT GT-3', CD14 tetra3-rev, respectively) in a conventional 25- μ l PCR with an extended elongation step (2.5 min 72°C) in a FlexCycler (Analytik Jena, Jena, Germany). The amplicons were then subjected either to 5'-nuclease genotyping assays for rs2569190 or rs4914 (see above) or to allele-specific amplification using tetra-primers for rs2563298 (rs2563 tetra G for, rs2563 tetra T rev, rs2563 tetra for, rs2563 tetra rev) in standard SYBR green reactions.

Quantification of CD14 mRNA

Quantification of gene expression was performed by real-time reverse transcriptase (RT)-PCR using the sequence detection system ABI prism 7000 (Applied Biosystems) as described [18]. Commercially available TaqMan gene

expression assays were used for determination of CD14 (Hs00169122g1), CD11b (Hs00355885m1), and GAPDH (Hs99999905m1) transcripts. To discriminate between transcript variant 1 and 2, the TaqMan genotyping assays for the SNPs rs2569190 and rs4914 were used in parallel.

Results

Analysis of allele-specific transcriptional activity in situ within a natural genomic context can be performed (a) on the level of gDNA determining serine 5 phosphorylated RNA polymerase II loading by HaploChIP technology as a parameter of transcription initiation and (b) on the level of mRNA determining the relative number of transcripts originating from two gene copies by ASTQ. Both techniques do not rely on the comparison of individuals but can be carried out with material from single individuals who are heterozygous at the position of interest. Whereas material from individuals who are heterozygous at a respective position is the only requirement for HaploChIP analyses, ASTQ requires material from individuals who are additionally heterozygous at a second SNP within the transcript (if the SNP of interest is located, e.g., in a nontranscribed regulatory region) and which must be used to discriminate between transcripts derived from one or the other sister chromosome. According to NCBI database, the CD14 transcript does contain a SNP suitable for ASTQ analysis, rs4914, due to its location (Fig. 1) and due to a sufficiently high MAF that should allow identification of double heterozygous individuals. Accordingly, from a total of 42 healthy individuals who were genotyped at positions rs2569190 and rs4914 (Table 1), 27 were found to be heterozygous at rs2569190, and seven were found to be heterozygous at both positions.

To cover all relevant SNPs within the CD14 gene and the nearby 5' and 3' regions, e.g., with a MAF >0.1, which might affect transcriptional activity, too, a panel of ten samples was sequenced spanning a region of 2,101 nucleotides (see Fig. 1a). Alignment of the sequences confirmed the four known SNPs rs5744455, rs2569190, rs4914, and rs2563298 and the absence of any further with

relevant MAFs (data not shown). Genotyping of rs5744455 and rs2563298 (Table 1) and haplotyping of the four respective loci was performed by a combination of ARMS-PCR and 5'-nuclease genotyping and revealed a total of five haplotypes (Table 2) resulting in a total of eleven haplotypes. C-T-C-G, the most frequent haplotype, was also the only one carrying a T allele at the position of interest, rs2569190.

HaploChIP analysis was performed on freshly prepared PBMC from five healthy rs2569190 heterozygous blood donors. Sheared chromatin from formalin-fixed cells served as the “input” control. Input chromatin and chromatin that was captured by an antibody directed against serine 5 phosphorylated RNA polymerase II, the “output” DNA, were analyzed for the ratio of fragments carrying the C or the T allele at position rs2569190 by RFLP analysis. Whereas input chromatin was found to contain about the same numbers of C and T variant fragments, the chromatin of all volunteers that had been captured by anti-RNA polymerase II antibodies was found to contain about twice the numbers of T variant fragments (Fig. 2a) suggesting that RNA polymerase II preferentially binds fragments containing the rs2569190 T allele. As an additional control, chromatin was also captured by a mock isotype (IgG) antibody. The mock captured material was found to contain nearly same amounts of C and T allele fragments (T/C ratios of 1.18, 1.02, and 1.21) as with the input control (T/C ratios 1.08, 1.10, and 1.11, respectively), while the output material was found to contain more T allele fragments (T/C ratios 1.90, 1.65, and 1.80, respectively; three independent samples, data not shown). An analysis of HaploChIP data with regard to haplogenotypes revealed the haplotype C-T-C-G that one to be preferentially captured by anti-RNA polymerase II antibodies, irrespectively of its genomic counterpart (Fig. 2b). As an additional control for validation of HaploChIP results, input and output chromatins were analyzed for the presence of two alleles of the SNP rs705 located within the imprinted gene SNRPN. Whereas input DNA contained similar amounts of each allele, immunoprecipitated material captured by antibodies directed against serine 5 phosphorylated RNA polymerase II contained only one allele as expected for the monoallelic expression (see Fig. 3b, lower panel).

Table 1 Genotype distribution among healthy blood donors

	Genotype distribution				
	wt/wt	wt/v	v/v	MAF	P ^b
rs5744455 [C/T] ^a	22 (52.4%)	18 (42.9%)	2 (4.7%)	0.26	0.7
rs2569190 [C/T] ^a	10 (23.8%)	27 (64.3%)	5 (11.9%)	0.44	0.07
rs4914 [C/G] ^a	31 (73.8%)	9 (21.4%)	2 (4.8%)	0.15	0.23
rs2563298 [G/T] ^a	22 (52.4%)	17 (40.5%)	3 (7.1%)	0.27	1

^a [wild-type (wt), variant (v)]

^b Exact test for Hardy–Weinberg equilibrium

820

J Mol Med (2009) 87:815–824

Table 2 Haplotype distribution among the healthy individuals under study

Haplotype ^a	Total number [n]
C-T-C-G	37
T-C-C-G	22
C-C-G-T	13
C-C-C-T	10
C-C-C-G	2

^a rs5744455–rs2569190–rs4914–rs2563298

To investigate whether a CD14 gene with a T allele at position rs2569190 is more actively transcribed than the C allele counterpart, ASTQ was performed on cDNA samples from double heterozygous donors using rs4914 for discrimination of transcripts. Discrimination of transcripts was achieved by two independent read-out assays: by RFLP analyses and by 5'-nuclease assays using differently labeled allele-specific probes. The amounts of C and G allelic transcript variants were not found to be different in all samples studied as they were neither, expectedly, within two samples from donors who are homozygous for the C allele at rs2569190. These results were achieved irrespectively of the technique applied (Fig. 3). Accordingly, analysis of ASTQ data with regard to haplogenotypes

revealed that the number of transcripts originating from C-T-C-G haplotypes does not exceed the number originating from the other haplotypes (Fig. 3a). As a control for allelic discrimination via RFLP analyses, cDNA samples from two homozygotes were mixed at various ratios and analyses yield expected results (Fig. 3a). As a control for allelic discrimination via allele-specific probes, gDNA and cDNA samples from individuals who were heterozygous at SNP rs705 within the imprinted gene SNRPN were analyzed for the ratio of the two alleles G and A. While in gDNA samples both alleles were detectable at nearly same amounts, corresponding cDNA samples contained either one or the other (Fig. 3b, lower panel).

The existence of two different CD14 mRNA variants, one of which carrying the SNP of interest within its 5'UTR (variant 1) and the other not at all (variant 2; Fig. 1b), might bias quantification of allelic transcript expression, e.g., by different half life times. To determine the relative proportion of transcript variants 1 and 2 within the cDNA samples, cDNA samples were analyzed for the total amount of CD14 transcripts on the one hand and for the amount of transcript variant 1 by using an assay spanning rs2569190. Whereas the number of transcripts containing rs4914 equals

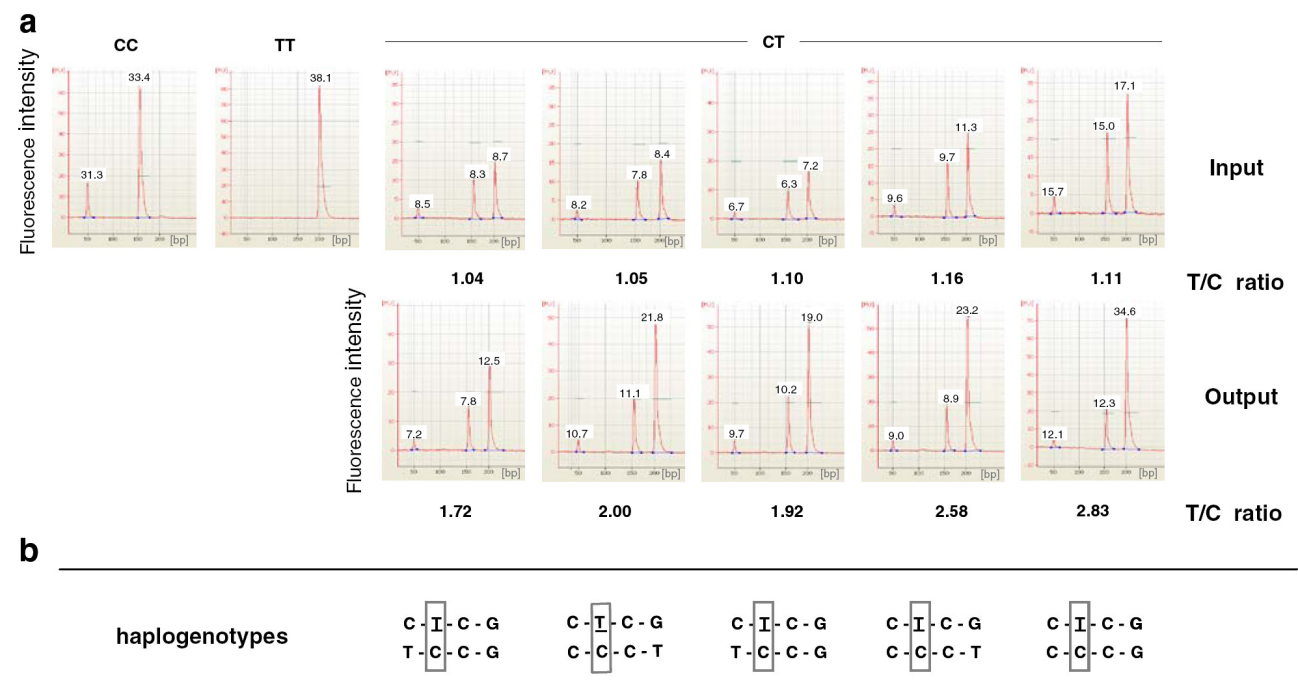


Fig. 2 Chromatin immunoprecipitation. **a** Formalin-fixed and enzymatically sheared chromatin from five healthy individuals heterozygous at rs2569190 was immunoprecipitated with antibodies directed against RNA polymerase II as described in the “Materials and methods” section or was left untreated as an “input” control. Purified input and output DNA samples were subjected to amplification spanning rs2569190. Products of 201 bp in size were subjected to restriction enzyme digestion with *HaeIII*. Restriction digest analysis

was carried out in a bioanalyzer yielding molar amounts of each fragment as given. Undigested fragments represent DNA of the T allele. Fragments of 49 and 152 bp in size represent DNA of the C allele. Material from homozygous individuals (TT and CC) is shown as a control for complete and specific enzymatic activity. **b** Haplotypes (rs5744455–rs2569190–rs4914–rs2563298) of the five donors are given

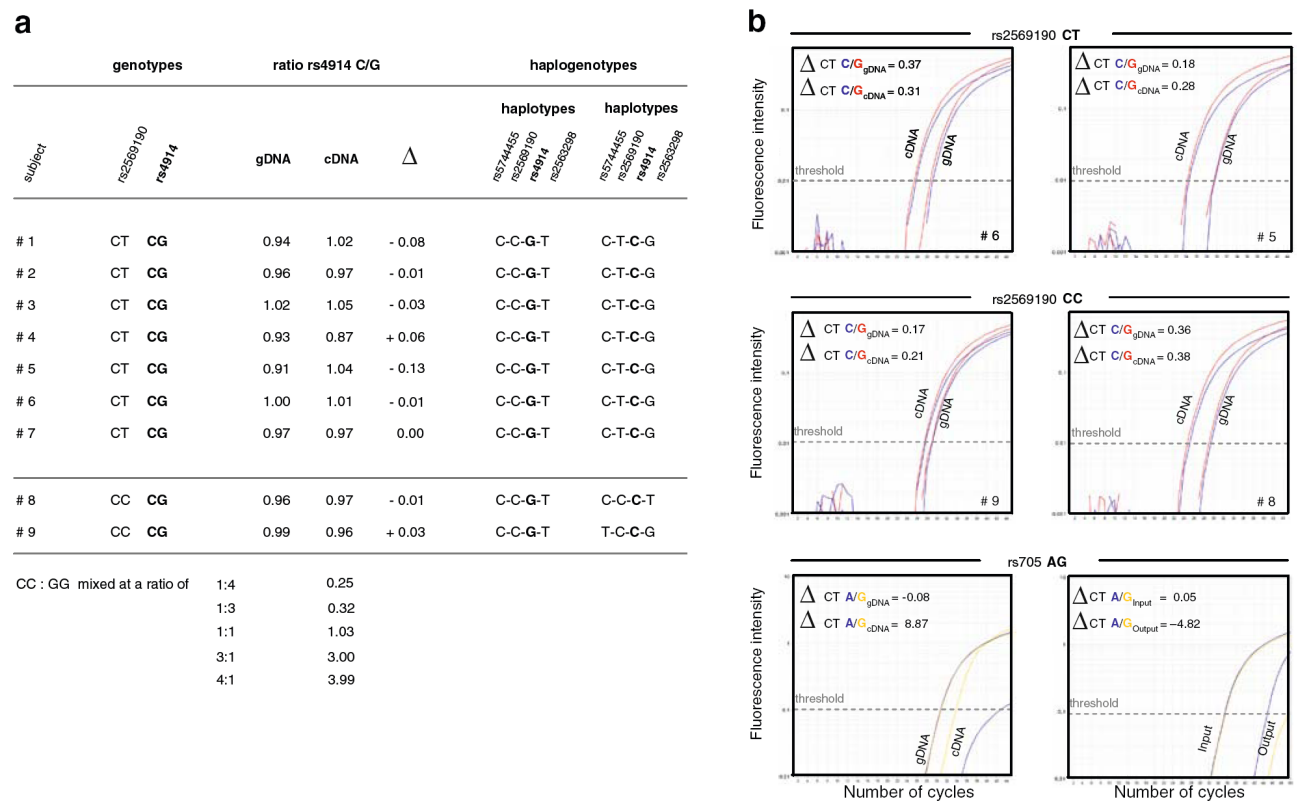


Fig. 3 ASTQ analyses. The relative amount of mRNA species carrying the C or the G allele at position rs4914 isolated from freshly prepared human PBMC samples was analyzed by RFLP and by 5'-nuclease assays using allele-specific probes. **a** RFLP data for all seven double heterozygous individuals are given as the ratio of rs4914 C/G alleles. Samples for two individuals homozygous at position rs2569190 (#8 and #9) as well as corresponding gDNA samples served as controls. cDNA samples that had been adjusted for the amount of CD14 mRNA by quantitative gene expression assays were

mixed at various ratios yielding expected C/G ratios. Haplotypes are indicated. **b** 5'-Nuclease assays were performed in triplicates. Data from one representative determination are given for two representative samples (#5 and #6; *top*), for two homozygous controls (#8 and #9; *middle*), and for a control of the imprinted gene SNPRN for genomic and cDNA and for HaploChIP material (*bottom*). Δ CT values for the rs4914 C (blue) and G alleles (red) in cDNA samples were not significantly different from the respective gDNA control

the number of total CD14 mRNA, the number of nucleic acids containing rs2569190 was about 100-fold lower and in the range of residual gDNA (Fig. 4). PBMC thus were found to contain nearly exclusively transcript splice variant 2. Analysis of total CD14 mRNA from freshly isolated PBMC failed to reveal a significant relationship with either the four single SNPs and with haplotypes when blood donors are divided into groups with none, one, or two C-T-C-G haplotypes (Fig. 5).

Discussion

According to earlier knowledge, the SNP rs2569190 was supposed to be located at position -159 within the promoter of the CD14 gene. On the basis of association studies, reporter gene assays, and EMSAs, this SNP was suggested to affect transcriptional activity by a differential binding of Sp1, Sp2, and Sp3 transcription factors [6, 8–10, 12].

Intriguingly, the impact of this SNP was shown to be different in monocytes and hepatoma cells, most probably due to a cell type specific different transcription factor pattern [12]. According to more recent knowledge, rs2569190 is located downstream of the transcription start site, and two splice variants do exist (Fig. 1). Its relative position within the (im-) mature transcripts depends on the splice variant; while rs2569190 is located within the 5' UTR of the transcript variant 1, this SNP is lacking in transcript variant 2 due to its intronic position.

PBMC were found to contain variant 2 CD14 transcripts (Fig. 4). Because of the absence of that SNP within the transcript, an effect of rs2569190 on transcript stability in PBMC can be ruled out. On the DNA level, the SNP might affect transcription factor binding as it is located in a transcription factor binding motif, a GC box, known to bind transcription factors Sp1, Sp2, and Sp3. LeVan et al. have shown both in EMSAs and in reporter gene assays that the T allele confers higher transcriptional activity in a mono-

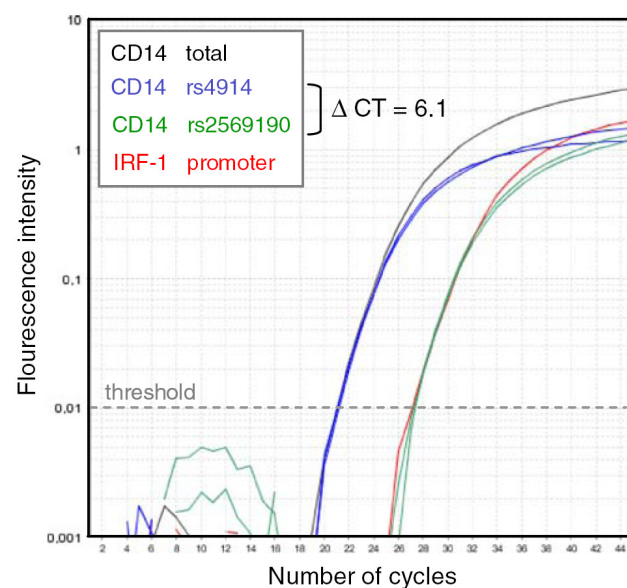


Fig. 4 Relative proportions of CD14 splice variants within PBMC cDNA samples. cDNA samples from seven double heterozygous individuals were analyzed for the amount of nucleic acids detected by an assay specific for total CD14 mRNA, by an assay specific for a region spanning rs4914, by an assay specific for a region spanning rs2569190, and by an assay specific for gDNA (IRF-1 promoter). The amount of products containing rs4914 equals the amount of total CD14 mRNA, whereas the amount of products containing rs2569190 equals that of gDNA

cytic cell line by preferential binding of the activating Sp1 and Sp2 transcription factors [12].

For the *in vivo* situation, i.e., in a natural genomic context, we found a stronger recruitment of RNA polymerase II to T allele variants. Transcription is known to follow a complex phosphorylation/dephosphorylation cycle in which RNA polymerase II is underlying a recycling process [15]. In brief, after assembly onto promoter sequences with general transcription factors (preinitiation), the RNA polymerase II core enzyme is phosphorylated on serines at position 5 by CDK7. Later on, after initiation of transcription, elongation requires phosphorylation on serines at position 2 by CDK9, a subunit of a transcription elongation factor. Polyadenylation signals transcription to terminate, to relieve RNA polymerase

II off its template, and finally to be dephosphorylated. HaploChIP analyses were performed using an antibody directed against serine 5 phosphorylated RNA polymerase II. Our data thus indicate that in the *in vivo* situation, the T allele confers stronger transcription initiation than the wild-type allele. Given that transcription factor binding might

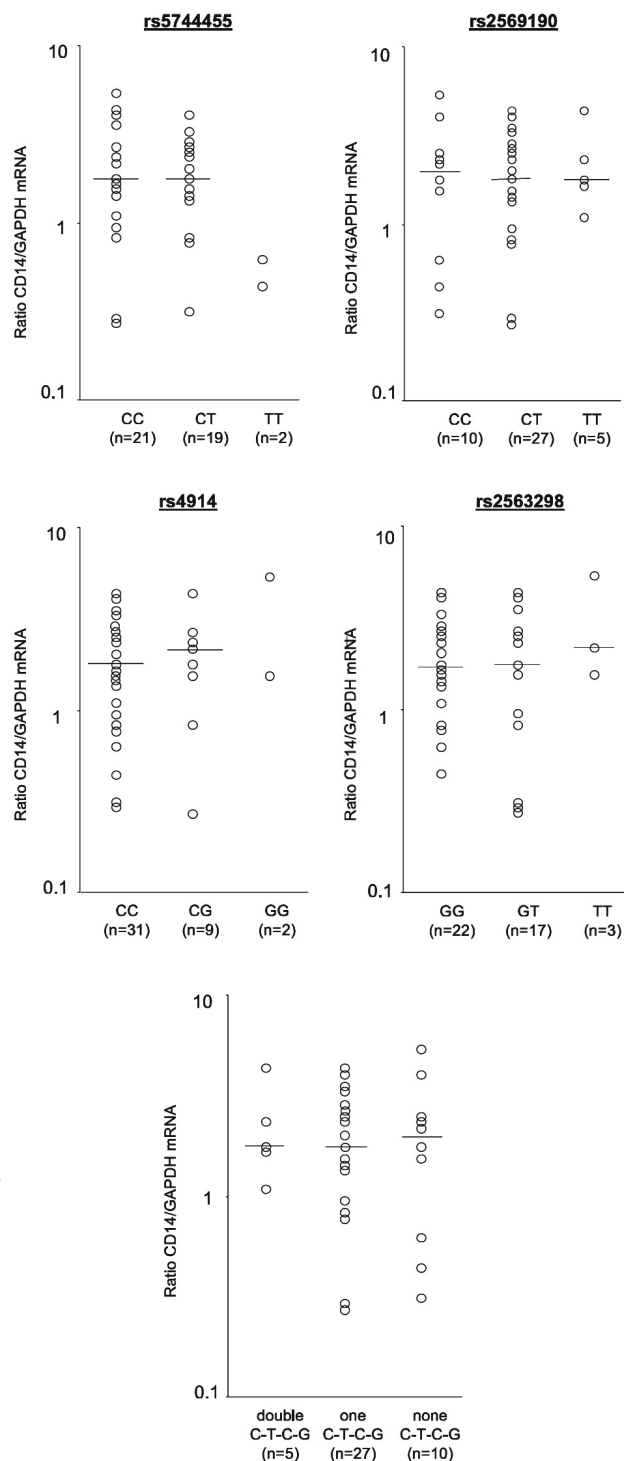


Fig. 5 CD14 mRNA expression in healthy donors with regard to their (haplo-)genotypes at four variant positions. Total CD14 mRNA was quantified in PBMC samples from 42 healthy blood donors using a commercially available gene expression assay by real time RT-PCR. Data were analyzed with regard to genotypes at the four respective positions and with regard to haplogenotypes as indicated. CD14 mRNA expression was not found to be different when donors with different (haplo-)genotypes were compared to each other, irrespectively of whether data were related to GAPDH as a reference transcript as shown or whether they were related to CD11b as a reference. Medians are given; for rs5744455 and rs4914, medians refer to wild-type homozygotes and variant allele carriers, respectively

stabilize RNA polymerase II binding, our data are thus in line with the findings by LeVan et al. [12].

By taking the number of transcripts as an end-point of transcription and by using two independent techniques for allelic discrimination, however, we found similar amounts of allelic mRNA variants. This finding is in accordance with data on total CD14 transcripts which do not differ in PBMC from donors with different genotypes (Fig. 5). We thus suppose that the T allele, or the C-T-C-G haplotype, might be advantageous for transcription initiation but not necessarily relevant for CD14 mRNA expression at least in PBMC.

While the rs2569190 T allele gene copy has been demonstrated to be preferentially loaded by RNA polymerase II in PBMC, endotoxin sensitivity appears not to be related to allele-specific CD14 gene expression in PBMC. This finding is in line with others showing different effects of regulatory variants in different tissues (reviewed in [16]). As outlined in the introduction, endotoxin sensing is not only mediated by mCD14 expressing myeloid cells but also by sCD14 able to transfer LPS to TLR4/MD2 expressing nonmyeloid cells lacking mCD14. Moreover, circulating sCD14 is known to inhibit systemic inflammatory responses to LPS via its ability to remove LPS from mCD14 and to transfer it to plasma lipoproteins which mediate clearance [1]. Thus, the associations of rs2569190 TT genotype with higher sCD14 levels and with lower levels of IgE might reflect an effect of that SNP independent on myeloid mCD14 but dependent on nonmyeloid, e.g., hepatic, sCD14 expression or action.

As sCD14 is described to have stimulatory effects in extravascular compartments, the association of rs2569190 T allele and progression in alcoholic liver disease might be also due to a better sensing of (endogenous) LPS from the gut via sCD14-mediated transfer of LPS to nonmyeloid liver cells. Based on the observation that hepatic stellate cells express TLR4 and respond to its stimulation, it has been suggested that TLR-mediated direct stellate cell activation is a major pathway to induce liver fibrosis [20, 21].

Several lines of evidence indicate that hepatocytes are the source of sCD14 [22–24]. To address the impact of rs2569190 on hepatic transcriptional activity of the CD14 gene, ASTQ analyses in liver tissue samples, e.g., from patients with a diagnosis of absence of liver disease, are required.

Acknowledgments The authors thank Waltraut Kopp for expert technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft [grant number MI 474/1-1]. R.B. was supported by a research scholarship by the German Academic Exchange Service (DAAD). E.A. is grateful for her scholarship from Damascus University, Syria.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Kitchens RL, Thompson PA (2005) Modulatory effects of sCD14 and LBP on LPS–host cell interactions. *J Endotoxin Res* 11:225–229
2. Gioannini TL, Weiss JP (2007) Regulation of interactions of Gram-negative bacterial endotoxins with mammalian cells. *Immunol Res* 39:249–260
3. Arbour NC, Lorenz E, Schutte BC, Zabner J, Kline JN, Jones M, Frees K, Watt JL, Schwartz DA (2000) TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. *Nat Genet* 25:187–191
4. El-Omar EM, Ng MT, Hold GL (2008) Polymorphisms in Toll-like receptor genes and risk of cancer. *Oncogene* 27:244–252
5. LeVan TD, Michel O, Dentener M, Thorn J, Vertongen F, Beijer L, Martinez FD (2008) Association between CD14 polymorphisms and serum soluble CD14 levels: effect of atopy and endotoxin inhalation. *J Allergy Clin Immunol* 121:434–440 e431. Epub 2007 Oct 2018
6. Baldini M, Lohman IC, Halonen M, Erickson RP, Holt PG, Martinez FD (1999) A Polymorphism* in the 5' flanking region of the CD14 gene is associated with circulating soluble CD14 levels and with total serum immunoglobulin E. *Am J Respir Cell Mol Biol* 20:976–983
7. Virta M, Pessi T, Helminen M, Seiskari T, Kondrashova A, Knip M, Hyoty H, Hurme M (2008) Interaction between CD14–159C&T polymorphism and *Helicobacter pylori* is associated with serum total immunoglobulin E. *Clin Exp Allergy* 38:1929–1934
8. Meiler C, Muhlbauer M, Johann M, Hartmann A, Schnabl B, Wodarz N, Schmitz G, Scholmerich J, Hellerbrand C (2005) Different effects of a CD14 gene polymorphism on disease outcome in patients with alcoholic liver disease and chronic hepatitis C infection. *World J Gastroenterol* 11:6031–6037
9. Jarvelainen HA, Orpana A, Perola M, Savolainen VT, Karhunen PJ, Lindros KO (2001) Promoter polymorphism of the CD14 endotoxin receptor gene as a risk factor for alcoholic liver disease. *Hepatology* 33:1148–1153
10. Von Hahn T, Halangk J, Witt H, Neumann K, Muller T, Puhl G, Neuhaus P, Nickel R, Beuers U, Wiedenmann B, Berg T (2008) Relevance of endotoxin receptor CD14 and TLR4 gene variants in chronic liver disease. *Scand J Gastroenterol* 43:584–592
11. Campos J, Gonzalez-Quintela A, Quinteiro C, Gude F, Perez LF, Torre JA, Vidal C (2005) The -159C/T polymorphism in the promoter region of the CD14 gene is associated with advanced liver disease and higher serum levels of acute-phase proteins in heavy drinkers. *Alcohol Clin Exp Res* 29:1206–1213
12. LeVan TD, Bloom JW, Bailey TJ, Karp CL, Halonen M, Martinez FD, Vercelli D (2001) A common single nucleotide polymorphism in the CD14 promoter decreases the affinity of Sp protein binding and enhances transcriptional activity. *J Immunol* 167:5838–5844
13. Buckland PR (2006) The importance and identification of regulatory polymorphisms and their mechanisms of action. *Biochim Biophys Acta* 1762:17–28
14. Charles Knight J (2005) HaploChIP: an in vivo assay. *Methods Mol Biol* 311:49–60
15. Palancade B, Bensaude O (2003) Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation. *Eur J Biochem* 270:3859–3870
16. Pastinen T, Hudson TJ (2004) Cis-acting regulatory variation in the human genome. *Science* 306:647–650
17. Knight JC, Keating BJ, Rockett KA, Kwiatkowski DP (2003) In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. *Nat Genet* 33:469–475 Epub 2003 Mar 2010

18. Mihm S, Frese M, Meier V, Wietzke-Braun P, Scharf JG, Bartenschlager R, Ramadori G (2004) Interferon type I gene expression in chronic hepatitis C. *Lab Invest* 84:1148–1159
19. Ye S, Dhillon S, Ke X, Collins AR, Day IN (2001) An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Res* 29:E88
20. Seki E, De Minicis S, Osterreicher CH, Kluwe J, Osawa Y, Brenner DA, Schwabe RF (2007) TLR4 enhances TGF-beta signaling and hepatic fibrosis. *Nat Med* 13:1324–1332
21. Paik YH, Schwabe RF, Bataller R, Russo MP, Jobin C, Brenner DA (2003) Toll-like receptor 4 mediates inflammatory signaling by bacterial lipopolysaccharide in human hepatic stellate cells. *Hepatology* 37:1043–1055
22. Bas S, Gauthier BR, Spenato U, Stingelin S, Gabay C (2004) CD14 is an acute-phase protein. *J Immunol* 172:4470–4479
23. Pan Z, Zhou L, Hetherington CJ, Zhang DE (2000) Hepatocytes contribute to soluble CD14 production, and CD14 expression is differentially regulated in hepatocytes and monocytes. *J Biol Chem* 275:36430–36435
24. Su GL, Dorko K, Strom SC, Nussler AK, Wang SC (1999) CD14 expression and production by human hepatocytes. *J Hepatol* 31:435–442

6.2. Functional relevance of the *IRF-1* promoter polymorphism rs2549009 on transcriptional activity in a native genomic environment

Jasmin Mertens, Giuliano Ramadori, Sabine Mihm

Department of Gastroenterology and Endocrinology, University Medical Center Goettingen,

Georg-August-Universitaet, D-37075 Goettingen, Germany

Published in: Human Molecular Genetics (2010) 19(23): 4587-4594

DOI: 10.1093/hmg/ddq386

Author contribution:

This study was designed and supervised by SM and GR. HaploChIP experiments were performed by JM. Data were analysed and manuscript was written jointly by JM and SM.

Functional relevance of the *IRF-1* promoter polymorphism rs2549009 on transcriptional activity in a native genomic environment

Jasmin Mertens, Giuliano Ramadori and Sabine Mihm*

Department of Gastroenterology and Endocrinology, University Medical Center Goettingen, Georg-August-Universitaet, D-37075 Goettingen, Germany

Received July 14, 2010; Revised and Accepted September 3, 2010

Interferon regulatory factor-1 (IRF-1), a transcription regulator involved both in inducing and in mediating the effects of interferon, is encoded by a highly polymorphic gene in different ethnic populations. Some of these genetic variations have been described to be associated to disease traits in hepatitis C virus and in human immunodeficiency virus infection, including one single-nucleotide polymorphism rs2549009 within the promoter region. This study aimed at investigating the functional relevance of rs2549009 on *IRF-1* transcriptional activity in peripheral blood mononuclear cells in its natural genomic environment. Haplotype-specific chromatin immunoprecipitation using antibodies directed against both the transcriptionally inactive and active RNA polymerase II (RNAPII) and allele-specific transcript quantification techniques were applied to *ex vivo*-derived samples from healthy heterozygous donors. Inactive serine 5 phosphorylated RNAPII was found to be preferentially bound to the rs2549009 A allele in all donors investigated. Active serine 2 phosphorylated (ser2-P) RNAPII, in contrast, was found to be precipitable, depending on the donor, preferentially either with the A or the G promoter variants or without any preference. The ratio of rs2549009 A/G promoter variants engaged by ser2-P RNAPII was closely related to the relative frequency of the respective *IRF-1* transcripts, and relative allelic expression was found to be associated to total *IRF-1* gene expression. These results provide evidence for a bidirectional *IRF-1* gene expression imbalance that appears not to be solely controlled by rs2549009 in *cis* and may rely on a yet unidentified variant or haplotype or on environmental control in *trans*.

INTRODUCTION

Host genetic factors are supposed to play a major role in determining differential susceptibility to human infectious diseases, in particular, in genes that are involved in innate immunity (1–3). One example are variations in the gene encoding for interferon (IFN) regulatory factor-1 (IRF-1), a transcription factor involved both in Toll-like receptor signalling and in mediating the effects of type I IFN (4,5). Moreover, IRF-1 initiates IFN- γ -mediated major histocompatibility complex class I antigen presentation-associated gene expression (6). Along with various levels of *IRF-1* expression regulation (5), genetic associations with susceptibility to human immunodeficiency virus-1 (HIV-1) and hepatitis C virus (HCV) infection are described for single-nucleotide polymorphisms

(SNPs) within introns 1 and 9 and a microsatellite region within intron 7 (7,8) or further intronic SNPs and a SNP at position –300 within the promoter region (9–11), respectively (Fig. 1).

Despite convincing associations of single variations with disease traits, functionality of the single variation needs to be shown (12,13). A functional or regulatory SNP is defined by conferring a distinct phenotype. A functional SNP within the promoter region of a gene, for instance, might affect its transcriptional activity (12,13). A demanding task in the field is still how to identify regulatory SNPs within gene regulatory regions. Reporter gene assays is one approach to study the impact of allelic promoter variants on reporter gene transcription within cells that have been transfected with a respective construct. Techniques that allow assessing transcriptional

*To whom correspondence should be addressed at: Department of Gastroenterology and Endocrinology, University Medical Center Goettingen, Georg-August-Universitaet, Robert-Koch-Straße 40, 37075 Goettingen, Germany. Tel: +49 551398946; Fax: +49 551397855; Email: smihm@med.uni-goettingen.de

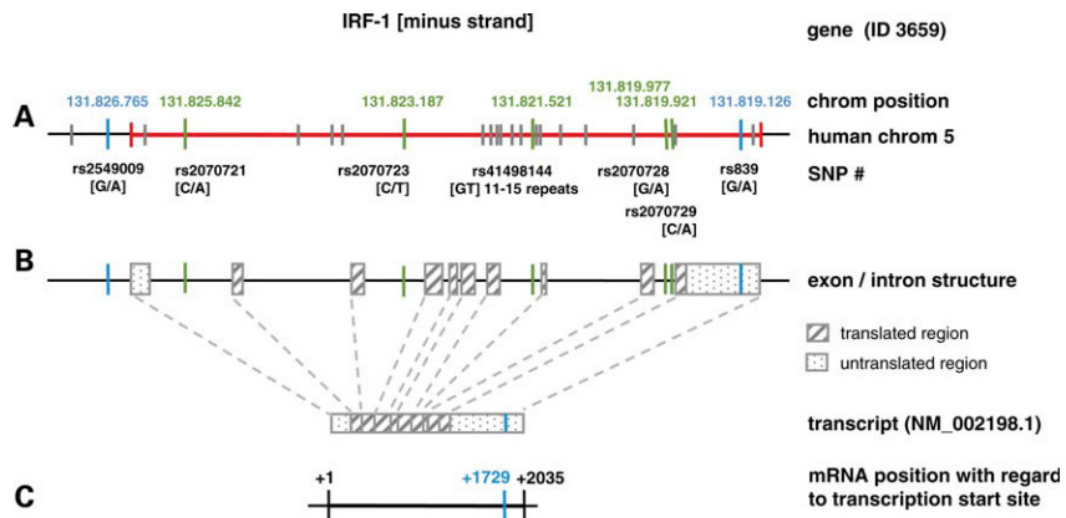


Figure 1. Schematic overview of the organization and variability of the *IRF-1* gene and transcript. In humans, the *IRF-1* gene is located on chromosome 5. The gene and the nearby 5' and 3' region include 25 variant positions (marked in grey) with a MAF > 0.1 known so far for the Caucasian population (NCBI database). rs2549009 and rs839 are indicated in blue within the chromosome (A), the gene (B) and the transcript (C). The positions of genetic variations that have been shown to be associated in addition with susceptibility to HIV-1 infection (rs2070721, rs41498144, rs2070729) or to HCV infection (rs2549009, rs2070723, rs2070728, rs2070729, rs839) are shown in green (7,11).

activity of allelic promoter variants in their natural genomic environment, however, have been developed more recently and might be superior: haplotype-specific chromatin immunoprecipitation (HaploChIP) using antibodies directed against RNA polymerase II (RNAPII) assesses the binding of the enzyme to allelic promoter variants as a surrogate for promoter activity. By using antibodies able to distinguish between serine 5- and serine 2-phosphorylated (ser5-P/ser2-P) heptapeptide repeats in the C-terminal domain (CTD) of the catalytic subunit of RNAPII, formation of a pre-initiation complex and transcript elongation can be discriminated as two stages of the transcription cycle, respectively (14,15). Allele-specific transcript quantification (ASTQ), as an additional approach, allows the quantification of mature transcripts that have been generated by the two parental promoter variants (16,17). By comparing two alleles within single heterozygous samples rather than by comparing groups of individual samples, these techniques are more sensitive in detecting *cis*-acting effects, and intrinsic errors that normally occur in inter-individual comparisons are reduced (17).

This study aimed at analysing whether SNP rs2549009, which is located within the *IRF-1* promoter region, confers any *cis*-regulatory effect on its transcriptional activity in a native genomic environment in freshly isolated peripheral blood mononuclear cells (PBMC) from heterozygous healthy blood donors by applying HaploChIP and ASTQ.

RESULTS

Determination of allele-specific promoter activity by HaploChIP

To assess *IRF-1* rs2549009 allele-specific promoter activity in a native sequence environment, engagement of A and G promoter variants by RNAPII was quantified by HaploChIP.

PBMC samples from 41 healthy Caucasian blood donors were genotyped at the respective position –300 within the promoter region of the *IRF-1* gene; 16 were found to be heterozygous and 7 to be homozygous for the variant allele A. The resulting minor allele frequency (MAF) of 0.366 was close to that given by the NCBI database and genotype distribution followed Hardy–Weinberg equilibrium (HWE) ($P = 0.322$, exact test).

To discriminate between the transcriptionally inactive and active isoforms of RNAPII, antibodies were chosen with specificity for the ser5-P and ser2-P enzymes, H14 and H5, respectively. Material from five heterozygous donors captured by these two antibodies, by a mock isotope antibody, and non-immunoprecipitated material (input) was analysed for the relative amount of *IRF-1* rs2549009 A and G allele DNA fragments by restriction fragment length polymorphism (RFLP). Non-immunoprecipitated input material and mock antibody precipitated controls expectedly showed nearly no difference in the relative amount of A and G allele DNA fragments. H14 immunoprecipitated material, however, was found to consistently contain about 1.4-fold higher amounts of A than of G allele fragments (Fig. 2A), implying that the transcriptionally inactive ser5-P RNAPII is preferentially bound to the rs2549009 A alleles in *ex vivo*-derived PBMC.

The ratio of rs2549009 A and G promoter variants in H5 captured DNA, in contrast, was found to vary reproducibly between 0.79 ± 0.01 (donor #1) and 1.24 ± 0.02 (donor #9) (Fig. 2), arguing for an allelic imbalance in both directions or, in other terms, against a preferential progression of RNAPII through the transcription cycle on a particular promoter allele variant.

Relative allelic expression of *IRF-1* transcripts

To address the question of whether allelic promoter activity is related to an allelic imbalance in mature *IRF-1* transcripts,

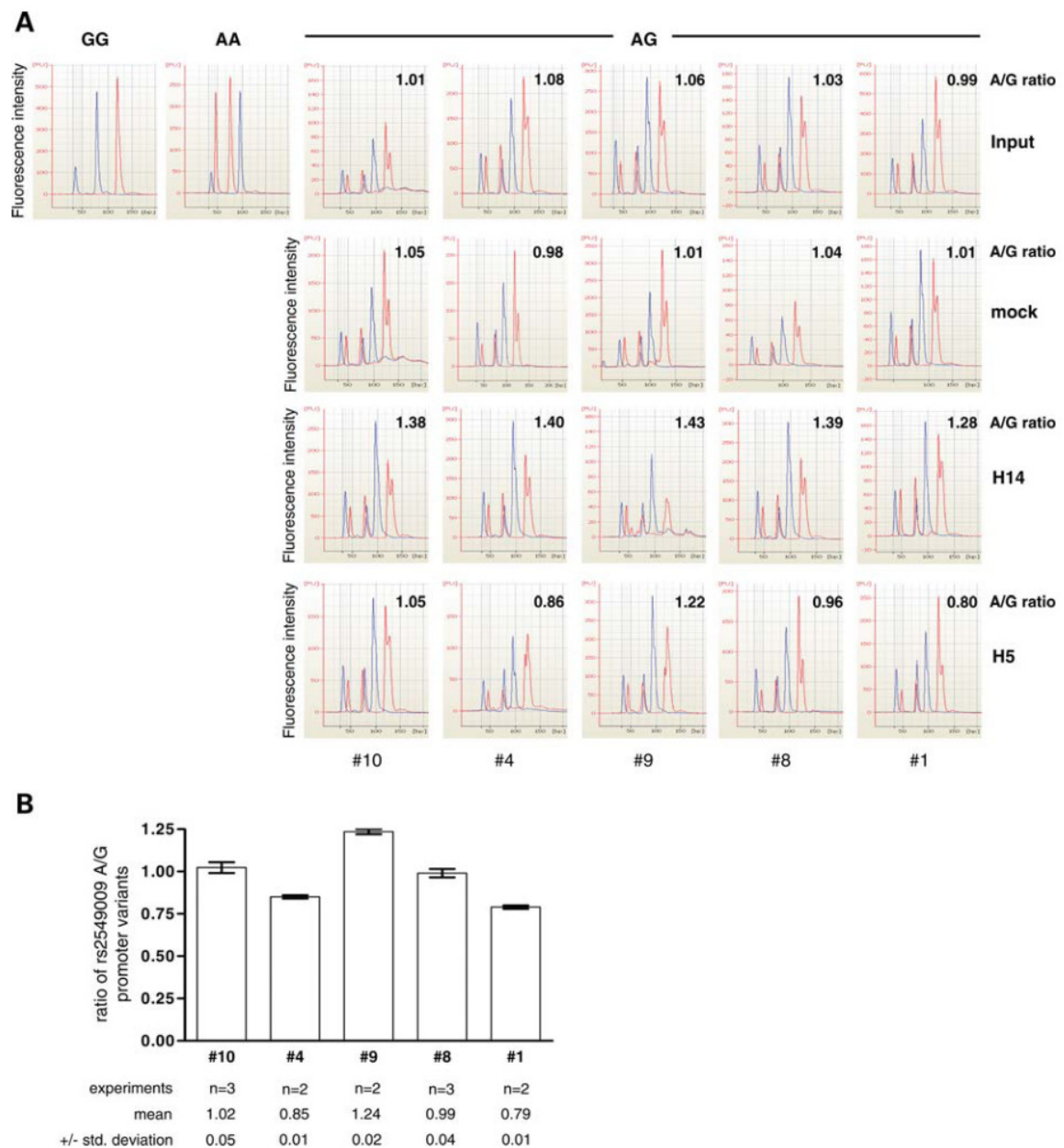


Figure 2. Chromatin immunoprecipitation. (A) Formalin-fixed and enzymatically sheared chromatin from five healthy individuals, who are heterozygous at SNP rs2549009 and rs839, was immunoprecipitated with antibodies directed against ser5-P and ser2-P RNAPII, H14 and H5, respectively, or a mock isotype. Non-immunoprecipitated material was used as an additional input control. Purified DNA fragments were subjected to amplification spanning a region around rs2549009 and yielding a product of 142 bp in size. Products were subjected to two different restriction enzyme digestions with *MvaI* and *BclI*. Digestion with *MvaI* (indicated in red) results in fragments of 69, 49 and 24 bp in size representing A allele promoter variant, while fragments of 118 and 24 bp represent G allele promoter variant. Digestion with *BclI* (indicated in blue) results in fragments of 91, 36 and 15 bp in size representing A allele, while fragments of 72, 36, 19 and 15 bp represent G allele promoter variants. Quantification of fragments was carried out in a bioanalyzer. Ratios of variant A and wild-type G allele promoter variants were calculated from the molar quantities of the respective fragments. Material from homozygous individuals (AA and GG) is shown as controls for complete and specific enzymatic activity. (B) Experiments were performed at least two times and mean ratios and standard deviations are given. In total, 12 independent immunoprecipitations were carried out on 9 independent cross-linkings of freshly isolated PBMC samples.

ASTQ assays were established. Aiming at identifying a SNP within the transcript suitable for assigning promoter to transcript variants, the cohort of 41 healthy blood donors was genotyped for the SNP rs839 located within *IRF-1* exon 10 as

well. Genotyping resulted in a distribution identical to the rs2549009 genotypes, thus being in line with HWE ($P = 0.322$, exact test), and a MAF of 0.366 close to that given by NCBI. Noteworthy, all individuals homozygous for the

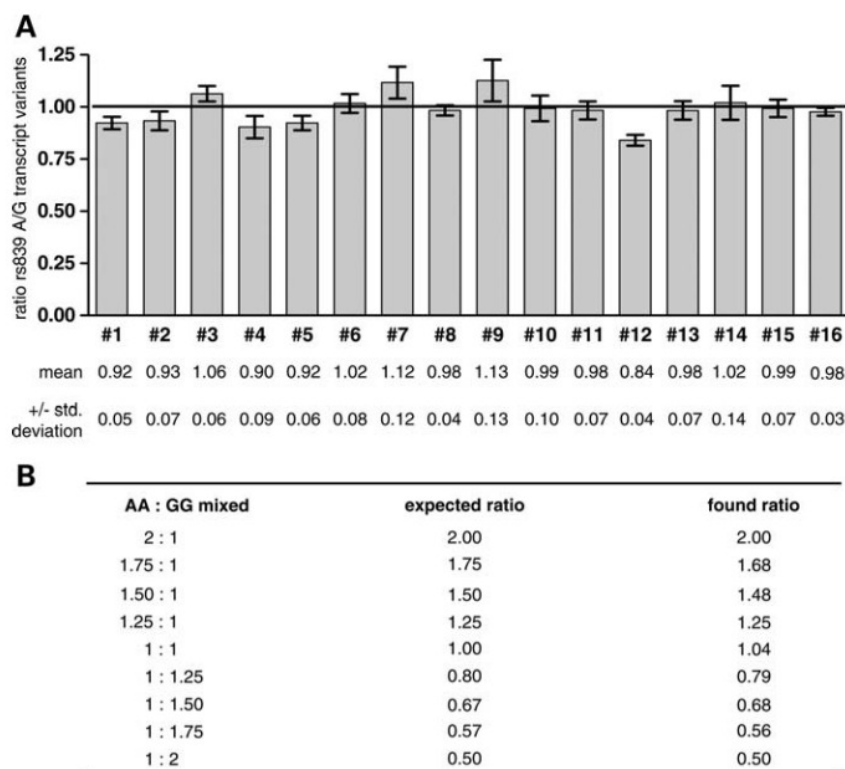


Figure 3. Allelic expression of IRF-1 transcripts. (A) The variant position rs839 within exon 10 was used to quantify mature IRF-1 transcripts and to discriminate transcript variants derived from the rs2549009 A or G promoter gene variant. A/G ratios of transcript variants in freshly prepared human PBMC samples from 16 individuals are shown. Discrimination and quantification of the two allelic variants was achieved by 5'-nuclease assays using allele-specific probes. Measurements were performed three times independently each in duplicates. Mean values and standard deviations are given. (B) Accuracy of allelic discrimination by allele-specific probes was assessed by mixing cDNA samples from G and A homozygous individuals at the given ratios and subjecting them to the read-out assay. Results closely met the expected ratios.

IRF-1 promoter rs2549009 wild-type allele G were found to be homozygous for the exon 10 rs839 wild-type allele G, too. With only two exceptions, the same was true for the variant allele A and for the heterozygous genotype findings. One rs2549009 heterozygous donor was found to be homozygous for the variant allele in rs839, while another donor homozygous for the A allele in rs2549009 showed a heterozygous genotype at rs839. On the basis of these findings, the degree of linkage disequilibrium (LD) was calculated. With a $D' = 0.947$ and $r^2 = 0.897$ the wild-type alleles G and the variant alleles A for both SNPs were shown to be in high and thus in suitable linkage for assigning transcript to the respective promoter variants.

The relative amounts of allelic transcripts in *ex vivo*-derived PBMC were quantified independently by two different read-out assays: (i) by 5'-nuclease assays using differently labelled allele-specific probes, a method that has been validated before using genomic DNA (gDNA) and cDNA samples from individuals heterozygous at a SNP within an imprinted gene (18) and (ii) by RFLP analyses using restriction endonucleases specific for each allele and quantification of fragments with a bioanalyzer. Among 16 individuals, the rs839 A/G ratios varied between 0.84 (donor #12) and 1.13 (donor #9) (Fig. 3A). Similar results were obtained by using RFLP analyses as a read-out with rs839 A/G ratios varying

between 0.87 and 1.06. A linear regression analysis of the data derived from the two different read-out assays revealed a close correlation ($P = 0.003$, $r = 0.702$) (data not shown).

The relative engagement of *IRF-1* rs2549009 A and G promoter variants with ser2-P RNAPII correlates to the relative allelic expression of IRF-1 transcripts

The relative proportion of allelic DNA fragments engaged by transcriptionally active RNAPII was found to be related to respective allele-specific *IRF-1* gene expression: those individuals who showed a preferential occupation of the rs2549009 A or G promoter variant by the active enzyme did reveal higher amounts of rs839 A or G IRF-1 transcript variants, respectively. In detail, PBMC obtained from donor #9 had an rs2549009 A/G promoter variant ratio of 1.22 and an rs839 A/G transcript ratio of 1.13 (Figs 2 and 3). PBMC from donors #1 and #4 with a higher promoter activity on the rs2549009 G allele in terms of a preferential binding of the active RNAPII, i.e. A/G ratios < 1 , also exhibited an rs839 A/G transcript ratio < 1 . Furthermore, samples from donors #8 and #10 with nearly equal activity on both rs2549009 promoter variants failed to exert any imbalance in allelic transcript expression (A/G ratios 0.98 and 0.99) (Figs 2 and 3). A linear regression analysis revealed a close

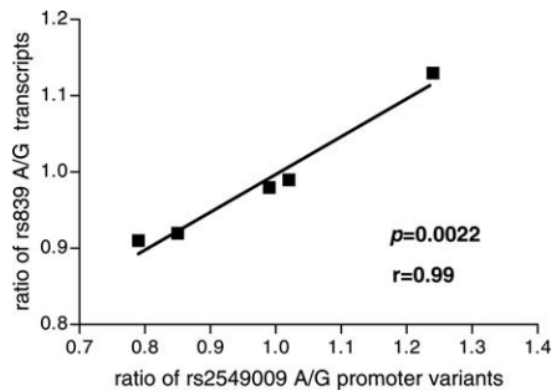


Figure 4. The relative binding of transcriptionally active RNAPII to *IRF-1* gene promoter variants correlates to the relative allelic expression of *IRF-1* transcripts. The relative frequency of rs2549009 promoter variant fragments in H5 captured material was related to the ratio of *IRF-1* rs839 A and G transcript variants in freshly isolated PBMC samples from five double-heterozygous individuals. A linear regression analysis revealed a close correlation of the findings obtained by the two *in vivo* approaches.

correlation between the presence of transcriptionally active RNAPII on the particular promoter variant and the amount of respective allele-specific *IRF-1* transcripts (Fig. 4).

Relative allelic *IRF-1* expression is associated to total *IRF-1* gene expression

In addition to the assessment of relative allelic *IRF-1* expression in rs2549009 heterozygotes, total *IRF-1* gene expression was quantified in freshly isolated PBMC samples as well. Data were related to glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as a reference transcript. Within the group of rs2549009 heterozygous donors, total *IRF-1* mRNA levels were found to be significantly and positively related to the relative amount of G allele transcripts (Fig. 5A). A comparison of the levels of total *IRF-1* mRNA between genotypic groups, in contrast, did not reveal a significant difference ($P = 0.2709$) (Fig. 5B).

DISCUSSION

HaploChIP and ASTQ are two techniques that enable studying the impact of genetic variants in their natural genomic context by using *ex vivo*-derived material. Applying HaploChIP by using antibodies directed against RNAPII, in particular those that are able to discriminate between phosphorylated isoforms (14,15,19–21), the combination of both techniques is expected to allow valuable insights into the activity of allelic promoter variants.

In PBMC from healthy blood donors, ser5-P RNAPII was consistently found to be preferentially recruited to the rs2549009 A allele promoter variant indicating a stronger transcription initiation activity on the variant A than on the wild-type G promoter (Fig. 2). Progression through the transcription cycle in terms of RNAPII serine 2 phosphorylation, however, was either more pronounced to the A or the G alleles or without any preference, depending on the donor (Fig. 2).

Ser5-P RNAPII has been shown to preassemble to the promoters of many mammalian genes across the genome in the absence of gene induction (22,23). The pre-assembled enzyme, however, fails to generate mature protein-coding transcripts (24). Induction of gene expression was shown to be controlled at the transition from initiation to elongation by the recruitment of kinases and subsequent phosphorylation of serine 2 residues within the CTD of RNAPII (24). We suppose that the variant A allele might be advantageous for pre-assembly and transcription initiation, but that it does not necessarily impact the transition towards transcriptional elongation in PBMC. The observation of a donor-dependent allele-specific ser5-P RNAPII activity is compatible with *in vitro* findings on *IRF-1* promoter activity showing stronger promoter activity either by the one or the other allele depending on the system (25,26). Whereas Saito *et al.* (25) were able to show a stronger activity of the A allele promoter variant in unstimulated or IFN-stimulated PLC/PRF/5 hepatoma cells, Schedel *et al.* (26) found a higher activity of a G allele promoter construct in ionomycin-stimulated Jurkat T cells.

Data on promoter activity as assessed by HaploChIP were complemented by ASTQ assessing the relative number of allelic mature transcripts originating from *IRF-1* rs2549009 promoter variants (Fig. 3). An allelic imbalance could be demonstrated in samples from those individuals in whom a preferential engagement of the respective promoter variants by ser2-P isoform of RNAPII was observed (Fig. 4).

This bidirectional imbalance in allele-specific transcription in rs2549009 heterozygous samples argues against a sole *cis*-regulatory role of rs2549009 in *IRF-1* gene expression. *IRF-1* gene expression might be regulated either by other *cis*-acting elements as well or it might be subject to gene–environment interactions in *trans*. The *IRF-1* rs2549009 SNP is in close neighbourhood to two other SNPs, rs2549007 and rs2706384, which also have been identified by Saito *et al.* (10) and which are located 110 and 115 nt upstream of rs2549009, respectively. The high degree of LD that has been found for the Japanese population is true for our cohort of healthy Caucasians for rs2549007, too ($D' = 1.0$, $r^2 = 0.947$) (unpublished data). Because of this high degree of LD, any observation made for rs2549009 might be attributable as well to rs2549007. Regarding rs2706384, MAF is lower in our Caucasian cohort, but allelic LD is close ($D' = 0.926$, $r^2 = 0.604$) (unpublished data). Thus, any observation made for rs2549009 might be attributable also to rs2706384, and haplotypes might have to be regarded.

An impact of environmental factors on rs2549009-dependent *IRF-1* gene expression is supported by a comparison of *IRF-1* gene expression and its inducibility by IFN- α in PBMC from healthy individuals and patients with chronic HCV infection. *IRF-1* mRNA expression levels in *ex vivo*-derived PBMC from hepatitis C patients were significantly higher than in healthy individuals, but neither significantly different with regard to rs2549009 genotype (27). Responsiveness of PBMC to human recombinant IFN- α *in vitro*, however, was significantly higher in healthy individuals, and, among healthy individuals, more pronounced in A allele carriers than in GG homozygous (unpublished data). The finding that relative allelic *IRF-1* expression is related to total *IRF-1* expression suggests, however, a biological impact of

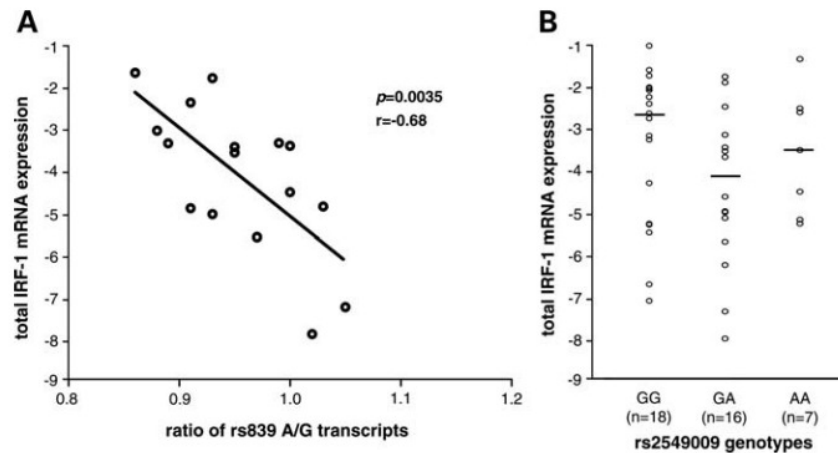


Figure 5. Association between relative allelic and total *IRF-1* gene expression. (A) Allele-specific *IRF-1* transcription was related to total *IRF-1* gene expression in PBMC of 16 heterozygous Caucasian blood donors. A close positive correlation was found between the relative proportion of G allele transcripts and total *IRF-1* mRNA. (B) The relative expression of total *IRF-1* transcripts in PBMC from 41 healthy Caucasian donors is given with regard to *IRF-1* rs2549009 genotypes. Data were related to GAPDH transcripts as a reference and are given as ΔCT . Medians are indicated. Statistical analysis (Kruskal–Wallis test) failed to reveal any significant difference in total *IRF-1* mRNA amounts between genotypic groups.

rs2549009 on total *IRF-1* gene expression. It is conceivable that the expression of either the rs2549009 A or the G allele gene is subject to a negative or a positive regulation in *trans*, respectively.

Allele-specific gene expression has been shown for up to 5% of human autosomal genes (28). Whole genome linkage studies of total expression have suggested that 1–3% of human genes harbour common *cis*-acting variants (29–31). Functional studies on single SNPs confirmed their impact on gene expression (32–34). Quite in accordance with our data, it also has been reported that imbalances of a particular gene may only occur in some individuals (30,31). The range of allelic imbalance can vary a lot not only between genes but also between individuals from less than 10% to 4.3-fold or higher (29,30,32,35).

The finding that allele-specific *IRF-1* gene transcription is associated to total *IRF-1* gene expression is evident only in heterozygous samples rather than becoming evident in comparing genotypic groups underscores the higher sensitivity of the *in vivo* approach for detecting functional *cis*-acting variations. Both, in HaploChIP and in ASTQ analyses, allelic promoter activity is assessed in heterozygote individual samples in a natural genomic context regardless of any *trans*-acting effects. However, given a sufficient enough number of samples, such a *cis*-acting mechanism might become apparent in the comparison of genotypic groups as well.

Interestingly, Ball *et al.* (7) reported on an HIV-1-resistant phenotype correlated with *IRF-1* gene polymorphisms and low *IRF-1* expression. Two of the variations linked to the HIV-1 protective phenotype (rs2070721, rs2070729) have also been associated to susceptibility towards HCV infection by Fortunato *et al.* (11). The contribution of SNPs that are located within the promoter region of a gene to transcriptional regulation is more obvious than a contribution of a SNP which is located within an intron, for instance. Some introns, however, are known to contain enhancer elements, while others elevate mRNA accumulation by a different process

that has been termed intron-mediated enhancement (IME). Although the mechanism remains to be unravelled, a model of IME has been proposed in which introns increase transcript elongation (36). Functional analyses of further, e.g. intronic, *IRF-1* gene variations together with approaches considering gene–environment interactions will enhance the understanding of the role of genetic variations for *IRF-1* gene expression and their relevance for individual susceptibility to viral infections.

MATERIALS AND METHODS

Blood donors

From 41 healthy Caucasian blood donors (20 female, 21 male; mean age 31.8 years, range 23–57), samples of 30 ml venous blood were taken for genotyping and for isolation of total cellular RNA. Blood donors with appropriate genotypes were, randomly chosen and consecutively on different days, asked to give further blood samples of 30 ml for HaploChIP experiments up to three times for independent replications ($n = 5$). All volunteers gave a written informed consent and the study was approved by the local ethical committee and conformed to the ethical principles of the 2000 Declaration of Helsinki.

Preparation of PBMC and isolation of total cellular RNA

From approximately 30 ml of heparinized peripheral blood, PBMC were prepared by Ficoll density centrifugation using guanidinium isothiocyanate, and total cellular RNA was isolated by CsCl density gradient ultracentrifugation essentially as described (37).

Genotyping

gDNA was purified from whole blood samples using the QIAamp DNA Blood Mini Kit following the blood and

body fluid spin protocol (Qiagen, Hilden, Germany). Typically, genotyping was performed by using 2–4 ng of PBMC-derived gDNA in a total volume of 10 µl in 5'-nuclease real-time polymerase chain reaction (PCR) assays carried out in the sequence detection system ABI StepOnePlus according to the supplier's instructions (Applied Biosystems, Darmstadt, Germany).

Genotyping of rs2549009 was carried out by amplifying gDNA with 0.36 µM of primers each (forward 5'-GGA CAA GGC GGA GTG AGA-3', reverse 5'-GGC TGC CCG GGA CAG-3') using the TaqMan Genotyping Master Mix (Applied Biosystems). Allelic discrimination was achieved by adding 0.2 µM differentially fluorescence dye-labelled allele-specific minor groove binder probes (VIC 5'-CCC CCG GCC CCA G-3', FAM 5'-CCC CTG GCC CCA G-3'). Genotyping of rs839, the marker SNP for ASTQ, was performed by using the commercially available TaqMan genotyping assay C_2554641_10.

HaploChIP

Chromatin immunoprecipitation (ChIP) was conducted using the ChIP-IT Express kit (Active Motif, Rixensart, Belgium) according to the supplier's protocol as described (18). In brief, cross-linked and enzymatically sheared chromatin corresponding to 1×10^6 cells was incubated with 4 µg of antibodies directed against ser5-P (H14, Covance, California, USA) or ser2-P (H5, Covance) RNAPII (38) or an IgM-mock control antibody (Linaris, Wertheim-Bettingen, Germany) or was used as control 'input' DNA. Antibody-bound protein/DNA complexes were immunoprecipitated by 25 µl magnetic protein G-coated beads that have been pretreated with 5 µg of a bridging anti-mouse-IgM antibody (Jackson ImmunoResearch, West Grove, PA, USA) for 1 h at 4°C according to the supplier's instructions.

After reverse cross-linking and treatment with proteinase K, 0.4 µl of input DNA, 4 µl of anti-RNAPII captured DNA and 4 µl mock control DNA were amplified in a total volume of 50 µl by PCR using 0.30 µM of primers each (IRF rs2549 for 5'-CAG GAG GGT GAA AAG ATG GCC-3', and IRF rs2549 rev 5'-CGA GCG CTC CCA ATC CAC-3'). Cycling conditions were a 3 min denaturation at 94°C, 45 cycles of 30 s denaturation at 94°C, annealing at 60°C for 30 s and elongation at 72°C for 30 s each, followed by a final elongation step for 10 min at 72°C. Products (142 bp) were purified and concentrated using the PrepEase gel extraction kit (USB, Cleveland, OH, USA). RFLP analyses were performed by subjecting the amplicons to restriction endonucleases digestion by *MvaI* and *BcnI* (Fermentas, St Leon-Rot, Germany) for 3 h. *MvaI* cleaves A allele products into fragments of 69, 49 and 24 bp in size, and G allele products into fragments of 118 and 24 bp, whereas digestion by *BcnI* leads to fragments of 91, 36 and 15 bp in size for A allele products and to fragments of 72, 36, 19 and 15 bp in size for G allele amplicons, respectively. To stop enzymatic digestion, reactions were cleaned a second time using the PrepEase gel extraction kit (USB) following the PCR purification protocol. Fragments were analysed in the Agilent 2100 bioanalyzer using the DNA 1000 LabChip kit

(Agilent Technologies, Böblingen, Germany). The relative amounts of A and G allele products were calculated.

ASTQ

Allele-specific quantification of IRF-1 transcript variants using rs839 as a marker SNP was achieved by (i) 5'-nuclease assay using two differently fluorescence dye-labelled allele-specific probes for the SNP rs839 and by (ii) RFLP analyses using two different allele-specific restriction endonucleases. Corresponding gDNA from double heterozygous samples was used as a control for both read-out assays. For 5'-nuclease assay, 6.4 ng cDNA and 4 ng gDNA were amplified in real-time PCR using a commercially available TaqMan genotyping assay (C_2554641_10).

For RFLP analyses, a product of 325 bp in size was generated by conventional PCR (IRF exon 10 for 5'-AGC CCA GCT GCC TGG AGA-3', IRF exon10 rev 5'-GAA ATG TGG CAA GAT CCA CAC G-3') and subjected to digestion by *LweI* and *BseMI* (Fermentas) for 3 h. *BseMI* cleaves A allele products into fragments of 97 and 228 bp in size, whereas digestion by *LweI* leads to fragments of 93 and 232 bp in size for G allele products, respectively. Fragments were purified and concentrated using the PrepEase gel extraction kit (USB) and afterwards separated and analysed in the Agilent 2100 bioanalyzer using the DNA 1000 LabChip kit (Agilent Technologies). The relative amounts of A and G allele products were calculated. Validation of used cDNA for both assays was performed by discrimination at SNP rs705 within the imprinted gene SNRPN (16).

Quantification of IRF-1 mRNA

Quantification of *IRF-1* gene expression was performed by real-time RT-PCR using the sequence detection system StepOnePlus (Applied Biosystems) as described (37). Commercially available TaqMan gene expression assays were used for the determination of IRF-1 (Hs00233698m1) and GAPDH (Hs99999905m1) transcripts.

Statistical analysis

Total gene expression analyses of 41 healthy blood donors with regard to promoter SNP genotypes were determined using one-way ANOVA (Kruskal–Wallis test) by applying the software GraphPad Prism version 4.00 (GraphPad Software Inc., La Jolla, CA, USA). Correlation and linear regression analyses were also performed using GraphPad Prism version 4.00. *P*-values < 0.05 are considered significant.

The degree of LD, expressed as D' and r^2 , between observed and expected frequencies of rs2549009/rs839 haplotypes and exact test for HWE were assessed by HaploView version 4.1 (39).

ACKNOWLEDGEMENTS

The authors thank Waltraut Kopp for expert technical assistance.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the Deutsche Forschungsgemeinschaft (grant number MI 474/1-1).

REFERENCES

- Lazarus, R., Vercelli, D., Palmer, L.J., Klimecki, W.J., Silverman, E.K., Richter, B., Riva, A., Ramoni, M., Martinez, F.D., Weiss, S.T. *et al.* (2002) Single nucleotide polymorphisms in innate immunity genes: abundant variation and potential role in complex human disease. *Immunol. Rev.*, **190**, 9–25.
- Hill, A.V. (2006) Aspects of genetic susceptibility to human infectious diseases. *Annu. Rev. Genet.*, **40**, 469–486.
- Misch, E.A. and Hawn, T.R. (2008) Toll-like receptor polymorphisms and susceptibility to human disease. *Clin. Sci. (Lond.)*, **114**, 347–360.
- Taniguchi, T., Ogasawara, K., Takaoka, A. and Tanaka, N. (2001) IRF family of transcription factors as regulators of host defense. *Annu. Rev. Immunol.*, **19**, 623–655.
- Tamura, T., Yanai, H., Savitsky, D. and Taniguchi, T. (2008) The IRF family transcription factors in immunity and oncogenesis. *Annu. Rev. Immunol.*, **26**, 535–584.
- Zhou, F. (2009) Molecular mechanisms of IFN-gamma to up-regulate MHC class I antigen processing and presentation. *Int. Rev. Immunol.*, **28**, 239–260.
- Ball, T.B., Ji, H., Kimani, J., McLaren, P., Marlin, C., Hill, A.V. and Plummer, F.A. (2007) Polymorphisms in IRF-1 associated with resistance to HIV-1 infection in highly exposed uninfected Kenyan sex workers. *AIDS*, **21**, 1091–1101.
- Ji, H., Ball, T.B., Ao, Z., Kimani, J., Yao, X. and Plummer, F.A. (2010) Reduced HIV-1 long terminal repeat transcription in subjects with protective interferon regulatory factor-1 genotype: a potential mechanism mediating resistance to infection by HIV-1. *Scand. J. Infect. Dis.*, **42**, 389–394.
- Wietzke-Braun, P., Maouzi, A.B., Manhardt, L.B., Bickeboller, H., Ramadori, G. and Mihm, S. (2006) Interferon regulatory factor-1 promoter polymorphism and the outcome of hepatitis C virus infection. *Eur. J. Gastroenterol. Hepatol.*, **18**, 991–997.
- Saito, H., Tada, S., Wakabayashi, K., Nakamoto, N., Takahashi, M., Nakamura, M., Ebinuma, H. and Ishii, H. (2002) The detection of IRF-1 promoter polymorphisms and their possible contribution to T helper 1 response in chronic hepatitis C. *J. Interferon Cytokine Res.*, **22**, 693–700.
- Fortunato, G., Calcagno, G., Bresciamorra, V., Salvatore, E., Filla, A., Capone, S., Liguori, R., Borelli, S., Gentile, I., Borrelli, F. *et al.* (2008) Multiple sclerosis and hepatitis C virus infection are associated with single nucleotide polymorphisms in interferon pathway genes. *J. Interferon Cytokine Res.*, **28**, 141–152.
- Hudson, T.J. (2003) Wanted: regulatory SNPs. *Nat. Genet.*, **33**, 439–440.
- Buckland, P.R. (2006) The importance and identification of regulatory polymorphisms and their mechanisms of action. *Biochim. Biophys. Acta*, **1762**, 17–28.
- Venters, B.J. and Pugh, B.F. (2009) How eukaryotic genes are transcribed. *Crit. Rev. Biochem. Mol. Biol.*, **44**, 117–141.
- Buratoski, S. (2009) Progression through the RNA polymerase II CTD cycle. *Mol. Cell*, **36**, 541–546.
- Knight, J.C., Keating, B.J., Rockett, K.A. and Kwiatkowski, D.P. (2003) In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. *Nat. Genet.*, **33**, 469–475.
- Pastinen, T., Ge, B. and Hudson, T.J. (2006) Influence of human genome polymorphism on gene expression. *Hum. Mol. Genet.*, **15**(Review Issue 1), R9–R16.
- Mertens, J., Bregadze, R., Mansur, A., Askar, E., Bickeboller, H., Ramadori, G. and Mihm, S. (2009) Functional impact of endotoxin receptor CD14 polymorphisms on transcriptional activity. *J. Mol. Med.*, **87**, 815–824.
- Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E. and Eick, D. (2007) Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science*, **318**, 1780–1782.
- Egloff, S., O'Reilly, D., Chapman, R.D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D. and Murphy, S. (2007) Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science*, **318**, 1777–1779.
- Akhtar, M.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D. and Ansari, A.Z. (2009) TFIIF kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol. Cell*, **34**, 387–393.
- Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R. and Young, R.A. (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*, **130**, 77–88.
- Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J. and Adelman, K. (2007) RNA polymerase is poised for activation across the genome. *Nat. Genet.*, **39**, 1507–1511.
- Hargreaves, D.C., Horng, T. and Medzhitov, R. (2009) Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell*, **138**, 129–145.
- Saito, H., Tada, S., Ebinuma, H., Wakabayashi, K., Takagi, T., Saito, Y., Nakamoto, N., Kurita, S. and Ishii, H. (2001) Interferon regulatory factor 1 promoter polymorphism and response to type 1 interferon. *J. Cell. Biochem.*, **81**(Suppl. 36), 191–200.
- Schedel, M., Pinto, L.A., Schaub, B., Rosenstiel, P., Cherkasov, D., Cameron, L., Klopp, N., Illig, T., Vogelberg, C., Weiland, S.K. *et al.* (2008) IRF-1 gene variations influence IgE regulation and atopy. *Am. J. Respir. Crit. Care Med.*, **177**, 613–621.
- Aidery, P., Ramadori, G. and Mihm, S. (2007) Increased IRF-1 and IFN- α gene expression in chronic hepatitis C G-300A IRF-1 promoter polymorphism A allele carriers. *Hepatology*, **46**, 448A.
- Pastinen, T., Ge, B., Gurd, S., Gaudin, T., Dore, C., Lemire, M., Lepage, P., Harmsen, E. and Hudson, T.J. (2005) Mapping common regulatory variants to human haplotypes. *Hum. Mol. Genet.*, **14**, 3963–3971.
- Serre, D., Gurd, S., Ge, B., Sladek, R., Sinnett, D., Harmsen, E., Bibikova, M., Chudin, E., Barker, D.L., Dickinson, T. *et al.* (2008) Differential allelic expression in the human genome: a robust approach to identify genetic and epigenetic cis-acting mechanisms regulating gene expression. *PLoS Genet.*, **4**, e1000006.
- Buckland, P.R. (2004) Allele-specific gene expression differences in humans. *Hum. Mol. Genet.*, **13**(Review Issue 2), R255–R260.
- Rockman, M.V. and Wray, G.A. (2002) Abundant raw material for cis-regulatory evolution in humans. *Mol. Biol. Evol.*, **19**, 1991–2004.
- Southam, L., Rodriguez-Lopez, J., Wilkins, J.M., Pombo-Suarez, M., Snelling, S., Gomez-Reino, J.J., Chapman, K., Gonzalez, A. and Loughlin, J. (2007) An SNP in the 5'-UTR of GDF5 is associated with osteoarthritis susceptibility in Europeans and with in vivo differences in allelic expression in articular cartilage. *Hum. Mol. Genet.*, **16**, 2226–2232.
- Bray, N.J., Jehu, L., Moskvina, V., Buxbaum, J.D., Dracheva, S., Haroutunian, V., Williams, J., Buckland, P.R., Owen, M.J. and O'Donovan, M.C. (2004) Allelic expression of APOE in human brain: effects of epsilon status and promoter haplotypes. *Hum. Mol. Genet.*, **13**, 2885–2892.
- Fogarty, M.P., Xiao, R., Prokunina-Olsson, L., Scott, L.J. and Mohlke, K.L. (2010) Allelic expression imbalance at high-density lipoprotein cholesterol locus MMAB-MVK. *Hum. Mol. Genet.*, **19**, 1921–1929.
- Yan, H., Yuan, W., Velculescu, V.E., Vogelstein, B. and Kinzler, K.W. (2002) Allelic variation in human gene expression. *Science*, **297**, 1143.
- Rose, A.B. (2008) Intron-mediated regulation of gene expression. *Curr. Top. Microbiol. Immunol.*, **326**, 277–290.
- Mihm, S., Frese, M., Meier, V., Wietzke-Braun, P., Scharf, J.G., Bartenschlager, R. and Ramadori, G. (2004) Interferon type I gene expression in chronic hepatitis C. *Lab. Invest.*, **84**, 1148–1159.
- Palancade, B. and Bensaude, O. (2003) Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation. *Eur. J. Biochem.*, **270**, 3859–3870.
- Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.

6.3. TLR3 gene polymorphisms and liver disease manifestations in chronic hepatitis C

Eva Askar,¹ Rusudan Bregadze,¹ Jasmin Mertens,¹ Stefan Schweyer,² Albert Rosenberger,³
Giuliano Ramadori¹ and Sabine Mihm¹

¹Department of Gastroenterology and Endocrinology, University Medical Center,
Georg-August-Universität, Göttingen, Germany

²Department of Pathology, University Medical Center, Georg-August-Universität Göttingen, Germany

³Department of Genetic Epidemiology, University Medical Center,
Georg-August-Universität Göttingen, Germany

Published in: Journal of Medical Virology (2009) 81:1204–1211

DOI 10.1002/jmv.21491

Author contribution:

This study was supervised by SM and GR. EA performed all experiments, analyses, and she wrote the manuscript nearly independently. Statistical analyses were performed by AR, and the histopathological evaluation of liver biopsy specimens by SS. As team members, JM and RB were intimately involved in the everyday discussion of data.

TLR3 Gene Polymorphisms and Liver Disease Manifestations in Chronic Hepatitis C

Eva Askar,¹ Rusudan Bregadze,¹ Jasmin Mertens,¹ Stefan Schweyer,² Albert Rosenberger,³ Giuliano Ramadori,¹ and Sabine Mihi^{1*}

¹Department of Gastroenterology and Endocrinology, University Medical Center, Georg-August-Universität, Göttingen, Germany

²Department of Pathology, University Medical Center, Georg-August-Universität, Göttingen, Germany

³Department of Genetic Epidemiology, University Medical Center, Georg-August-Universität, Göttingen, Germany

Phenotypes of liver disease due to chronic hepatitis C virus (HCV) infection show a wide range of variations in terms of histological manifestations and the clinical outcome. Sensing of viral double-stranded RNA (dsRNA) by Toll-like receptor 3 (TLR3) is likely involved in early pathogen detection and the host response to viral infection. This study analyzed epidemiological and clinical data from a total of 137 patients with chronic HCV infection with regard to two polymorphic positions within the TLR3 gene: rs5743305 (T/A) is located within the promoter region and might affect transcriptional activity, rs3775291 (C/T) is a non-synonymous single nucleotide polymorphism (SNP) located within exon 4 and the variant receptor has been shown to be functionally impaired. TLR3 promoter and the exon 4 variations were not found to be associated with TLR3 gene expression in peripheral blood mononuclear cells (PBMCs). In the liver, however, a tendency of higher TLR3 gene expression was found for exon 4 TT genotypes. Both variations were not found to be associated with clinical parameters of chronic disease. On the other hand, an analysis of the TLR3 exon 4 genotype distribution with respect to HCV subtype revealed an absence of TT genotype among HCV subtype 1a infected individuals. This study thus failed to reveal any association of the two SNPs under investigation with clinical parameters of chronic hepatitis C. However, data argue for a functional relevance of the exon 4 SNP in terms of conferring a different susceptibility towards HCV subtype infection. **J. Med. Virol.** 81:1204–1211, 2009. © 2009 Wiley-Liss, Inc.

KEY WORDS: single nucleotide polymorphism (SNP); hepatitis C virus (HCV); Toll-like receptor (TLR); HCV genotypes; T-976A; L412F

INTRODUCTION

Hepatitis C virus (HCV) is an enveloped single-stranded plus-sense RNA virus that causes acute and often chronic hepatitis [Choo et al., 1989]. Currently, an estimated 3% of the world's population—about 210 million people—is infected with HCV [Shepard et al., 2005]. The natural outcome of infection ranges from silent to self-limited to chronic [Micallef et al., 2006]. The natural course of chronic disease is also highly variable and ranges from asymptomatic to mild disease to cirrhosis and hepatocellular carcinoma (HCC) and HCV-related co-diseases [Seeff, 2002]. The complete pattern of host responses to HCV infection is still unclear. However, the HCV genome encodes regions of extensive secondary double-stranded RNA (dsRNA) structure, and the viral replication cycle includes dsRNA as an intermediate; both entities are assumed to be sensed by pattern recognition receptors (PRRs) during infection [Gale and Foy, 2005].

PRRs comprise membrane Toll-like receptors (TLRs), which are either expressed on the cell surface or on endosomal–lysosomal membranes, and cytosolic receptors such as nucleotide-binding oligomerization domain (NOD)-like receptors and the RNA-helicase family [reviewed in Akira et al., 2006; Takeuchi and Akira, 2007; Seki and Brenner, 2008]. Viral dsRNA is a ligand of TLR3 which is expressed within the endosomal compartment of conventional dendritic cells (cDCs) [Alexopoulou et al., 2001; Matsumoto et al., 2004]. Upon phagocytosis of infectious material and signaling through TLR3, DCs become activated. In the liver,

Grant sponsor: Deutsche Forschungsgemeinschaft; Grant number: MI 474/1-1.

*Correspondence to: Sabine Mihi, Division of Gastroenterology and Endocrinology, Department of Internal Medicine, Georg-August-Universität, Robert-Koch-Straße 40, Göttingen 37075, Germany. E-mail: smihm@med.uni-goettingen.de

Accepted 3 March 2009

DOI 10.1002/jmv.21491

Published online in Wiley InterScience
(www.interscience.wiley.com)

6.4. Congress contributions

Bregadze R, **Mertens J**, Mansur A, Ramadori G, Mihm S (2008)

Untersuchungen zur Assoziation des rs2569190/C-159T Polymorphismus im Gen des Endotoxinrezeptors CD14 mit der transkriptionellen Aktivität bei Patienten mit chronischer Hepatitis C
Z Gastroenterol 46: 953

Mertens J, Bregadze R, Askar E, Ramadori G, Mihm S (2009)

Assessment of the relevance of a CD14 gene polymorphism on RNA polymerase II DNA binding by haplotype-specific immunoprecipitation (HaploChIP)
Z Gastroenterol 47: 155

Bregadze R, **Mertens J**, Askar E, Mansur A, Ramadori G, Mihm S (2009)

Hepatic Endotoxin receptor CD14 expression in chronic hepatitis C patients with regard to rs2569190/C-159T genetic variants
Z Gastroenterol 47: 149

Askar E, Bregadze R, **Mertens J**, Ramadori G, Mihm S (2009)

Lack of an association between CD14 gene variants and disease manifestations in chronic hepatitis C
Z Gastroenterol 47: 148

Mertens J, Ramadori G, Mihm S (2009)

Increased CD14 transcription initiation in T allele carriers of the rs2549190/C-159T promoter polymorphism in healthy individuals and chronic hepatitis C patients
Falk Workshop "Immunology and Liver Disease", 15-16. Oktober 2009, Hannover

7. References

- Aidery, P. (2010). Untersuchungen zur Funktion des Polymorphismus G-300A / rs2549009 im Promoterbereich des IRF-1-Gens mittels Allele-spezifischer Transkript-Quantifizierung bei gesunden Probanden und bei Patienten mit chronischer HCV-Infektion. Medizinische Dissertation *Göttingen 2010*.
- Akira, S., Uematsu, S., and Takeuchi, O. (2006). Pathogen recognition and innate immunity. *Cell* 124, 783-801.
- Alexopoulou, L., Holt, A.C., Medzhitov, R., and Flavell, R.A. (2001). Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3. *Nature* 413, 732-738.
- Antal-Szalmás, P. (2000). Evaluation of CD14 in host defence. *Eur J Clin Invest* 30, 167-179.
- Arbour, N.C., Lorenz, E., Schutte, B.C., Zabner, J., Kline, J.N., Jones, M., Frees, K., Watt, J.L., and Schwartz, D.A. (2000). TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. *Nat Genet* 25, 187-191.
- Armant, M.A., and Fenton, M.J. (2002). Toll-like receptors: a family of pattern-recognition receptors in mammals. *Genome Biol* 3, REVIEWS3011.
- Arroyo-Espliguero, R., Avanzas, P., Jeffery, S., and Kaski, J.C. (2004). CD14 and toll-like receptor 4: a link between infection and acute coronary events? *Heart* 90, 983-988.
- Askar, E., Ramadori, G., and Mihm, S. (2009). Endotoxin receptor CD14 gene variants and histological features in chronic HCV infection. *World J Gastroenterol* 15, 3884-3890.
- Awadalla, P., Eyre-Walker, A., and Smith, J.M. (1999). Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science* 286, 2524-2525.
- Baldini, M., Lohman, I.C., Halonen, M., Erickson, R.P., Holt, P.G., and Martinez, F.D. (1999). A Polymorphism* in the 5' flanking region of the CD14 gene is associated with circulating soluble CD14 levels and with total serum immunoglobulin E. *Am J Respir Cell Mol Biol* 20, 976-983.
- Ball, T.B., Ji, H., Kimani, J., McLaren, P., Marlin, C., Hill, A.V., and Plummer, F.A. (2007). Polymorphisms in IRF-1 associated with resistance to HIV-1 infection in highly exposed uninfected Kenyan sex workers. *AIDS* 21, 1091-1101.
- Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263-265.
- Beutler, B., Jiang, Z., Georgel, P., Crozat, K., Croker, B., Rutschmann, S., Du, X., and Hoebe, K. (2006). Genetic analysis of host resistance: Toll-like receptor signaling and immunity at large. *Annu Rev Immunol* 24, 353-389.
- Beutler, B.A. (2009). TLRs and innate immunity. *Blood* 113, 1399-1407.
- Bregadze, R. (2010). Untersuchungen zur Assoziation genetischer Polymorphismen im Gen des Endotoxinrezeptors CD14 mit der transkriptionellen Aktivität. Medizinische Dissertation *Göttingen 2010*.

- Buckland, P.R. (2004). Allele-specific gene expression differences in humans. *Hum Mol Genet* 13 Spec No 2, R255-260.
- Buckland, P.R. (2006). The importance and identification of regulatory polymorphisms and their mechanisms of action. *Biochim Biophys Acta* 1762, 17-28.
- Campos, J., Gonzalez-Quintela, A., Quinteiro, C., Gude, F., Perez, L.F., Torre, J.A., and Vidal, C. (2005). The -159C/T polymorphism in the promoter region of the CD14 gene is associated with advanced liver disease and higher serum levels of acute-phase proteins in heavy drinkers. *Alcohol Clin Exp Res* 29, 1206-1213.
- Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., and Eick, D. (2007). Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* 318, 1780-1782.
- El-Omar, E.M., Ng, M.T., and Hold, G.L. (2008). Polymorphisms in Toll-like receptor genes and risk of cancer. *Oncogene* 27, 244-252.
- Fortunato, G., Calcagno, G., Bresciamorra, V., Salvatore, E., Filla, A., Capone, S., Liguori, R., Borelli, S., Gentile, I., Borrelli, F., *et al.* (2008). Multiple sclerosis and hepatitis C virus infection are associated with single nucleotide polymorphisms in interferon pathway genes. *J Interferon Cytokine Res* 28, 141-152.
- Godowski, P.J. (2005). A smooth operator for LPS responses. *Nat Immunol* 6, 544-546.
- Hargreaves, D.C., Horng, T., and Medzhitov, R. (2009). Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell* 138, 129-145.
- Hedrick, P., and Kumar, S. (2001). Mutation and linkage disequilibrium in human mtDNA. *Eur J Hum Genet* 9, 969-972.
- Heil, F., Hemmi, H., Hochrein, H., Ampenberger, F., Kirschning, C., Akira, S., Lipford, G., Wagner, H., and Bauer, S. (2004). Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science* 303, 1526-1529.
- Honda, K., and Taniguchi, T. (2006). IRFs: master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors. *Nat Rev Immunol* 6, 644-658.
- Iwasaki, A., and Medzhitov, R. (2004). Toll-like receptor control of the adaptive immune responses. *Nat Immunol* 5, 987-995.
- Jarvelainen, H.A., Orpana, A., Perola, M., Savolainen, V.T., Karhunen, P.J., and Lindros, K.O. (2001). Promoter polymorphism of the CD14 endotoxin receptor gene as a risk factor for alcoholic liver disease. *Hepatology* 33, 1148-1153.
- Johnson, A.D., Wang, D., and Sadée, W. (2005). Polymorphisms affecting gene regulation and mRNA processing: broad implications for pharmacogenetics. *Pharmacol Ther* 106, 19-38.
- Kim, S., and Misra, A. (2007). SNP genotyping: technologies and biomedical applications. *Annu Rev Biomed Eng* 9, 289-320.

- Kitchens, R.L., and Thompson, P.A. (2005). Modulatory effects of sCD14 and LBP on LPS-host cell interactions. *J Endotoxin Res* 11, 225-229.
- Knight, J.C., Keating, B.J., Rockett, K.A., and Kwiatkowski, D.P. (2003). In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. *Nat Genet* 33, 469-475.
- Kroger, A., Koster, M., Schroeder, K., Hauser, H., and Mueller, P.P. (2002). Activities of IRF-1. *J Interferon Cytokine Res* 22, 5-14.
- Kubistova, Z., Mrazek, F., and Petrek, M. (2009). Polymorphisms of the immune response genes: selected biological, methodical and medical aspects. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* 153, 93-102.
- Lazarus, R., Vercelli, D., Palmer, L.J., Klimecki, W.J., Silverman, E.K., Richter, B., Riva, A., Ramoni, M., Martinez, F.D., Weiss, S.T., *et al.* (2002). Single nucleotide polymorphisms in innate immunity genes: abundant variation and potential role in complex human disease. *Immunol Rev* 190, 9-25.
- LeVan, T.D., Bloom, J.W., Bailey, T.J., Karp, C.L., Halonen, M., Martinez, F.D., and Vercelli, D. (2001). A common single nucleotide polymorphism in the CD14 promoter decreases the affinity of Sp protein binding and enhances transcriptional activity. *J Immunol* 167, 5838-5844.
- Levan, T.D., Michel, O., Dentener, M., Thorn, J., Vertongen, F., Beijer, L., and Martinez, F.D. (2008). Association between CD14 polymorphisms and serum soluble CD14 levels: effect of atopy and endotoxin inhalation. *J Allergy Clin Immunol* 121, 434-440 e431.
- Mansur, A. (2008). Überprüfung der Funktionalität des genetischen Polymorphismus rs2569190/T-159C im CD14-Gen mittels Allel-spezifischer Transkript-Quantifizierung bei gesunden Probanden und bei Patienten mit chronischer Hepatitis-C-Virus-Infektion. Medizinische Dissertation *Göttingen* 2008.
- Maouzi, A. (2007). Assoziationen genetischer Polymorphismen in Genen, deren Produkte die angeborene Immunität vermitteln, mit klinischen und viralen Parametern einer Hepatitis-C-Virus-Infektion. Medizinische Dissertation *Göttingen* 2007.
- Mayo, O. (2008). A century of Hardy-Weinberg equilibrium. *Twin Res Hum Genet* 11, 249-256.
- Medzhitov, R., and Janeway, C., Jr. (2000). Innate immunity. *N Engl J Med* 343, 338-344.
- Meiler, C., Muhlbauer, M., Johann, M., Hartmann, A., Schnabl, B., Wodarz, N., Schmitz, G., Scholmerich, J., and Hellerbrand, C. (2005). Different effects of a CD14 gene polymorphism on disease outcome in patients with alcoholic liver disease and chronic hepatitis C infection. *World J Gastroenterol* 11, 6031-6037.
- Mertens, J., Bregadze, R., Mansur, A., Askar, E., Bickeboller, H., Ramadori, G., and Mihm, S. (2009). Functional impact of endotoxin receptor CD14 polymorphisms on transcriptional activity. *J Mol Med* 87, 815-824.

- Mertens, J., Ramadori, G., and Mihm, S. (2010). Functional relevance of the IRF-1 promoter polymorphism rs2549009 on transcriptional activity in a native genomic environment. *Hum Mol Genet* 19, 4587-4594.
- Miller, S.I., Ernst, R.K., and Bader, M.W. (2005). LPS, TLR4 and infectious disease diversity. *Nat Rev Microbiol* 3, 36-46.
- Palancade, B., and Bensaude, O. (2003). Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation. *Eur J Biochem* 270, 3859-3870.
- Palsson-McDermott, E.M., and O'Neill, L.A. (2004). Signal transduction by the lipopolysaccharide receptor, Toll-like receptor-4. *Immunology* 113, 153-162.
- Pampin, S., and Rodriguez-Rey, J.C. (2007). Functional analysis of regulatory single-nucleotide polymorphisms. *Curr Opin Lipidol* 18, 194-198.
- Pan, Z., Zhou, L., Hetherington, C.J., and Zhang, D.E. (2000). Hepatocytes contribute to soluble CD14 production, and CD14 expression is differentially regulated in hepatocytes and monocytes. *J Biol Chem* 275, 36430-36435.
- Pastinen, T., Ge, B., Gurd, S., Gaudin, T., Dore, C., Lemire, M., Lepage, P., Harmsen, E., and Hudson, T.J. (2005). Mapping common regulatory variants to human haplotypes. *Hum Mol Genet* 14, 3963-3971.
- Pastinen, T., and Hudson, T.J. (2004). Cis-acting regulatory variation in the human genome. *Science* 306, 647-650.
- Prelich, G. (2002). RNA polymerase II carboxy-terminal domain kinases: emerging clues to their function. *Eukaryot Cell* 1, 153-162.
- Rockman, M.V., and Wray, G.A. (2002). Abundant raw material for cis-regulatory evolution in humans. *Mol Biol Evol* 19, 1991-2004.
- Rose, A.B. (2008). Intron-mediated regulation of gene expression. *Curr Top Microbiol Immunol* 326, 277-290.
- Saito, H., Tada, S., Ebinuma, H., Wakabayashi, K., Takagi, T., Saito, Y., Nakamoto, N., Kurita, S., and Ishii, H. (2001). Interferon regulatory factor 1 promoter polymorphism and response to type 1 interferon. *J Cell Biochem Suppl* 36, 191-200.
- Saito, H., Tada, S., Wakabayashi, K., Nakamoto, N., Takahashi, M., Nakamura, M., Ebinuma, H., and Ishii, H. (2002). The detection of IRF-1 promoter polymorphisms and their possible contribution to T helper 1 response in chronic hepatitis C. *J Interferon Cytokine Res* 22, 693-700.
- Schedel, M., Pinto, L.A., Schaub, B., Rosenstiel, P., Cherkasov, D., Cameron, L., Klopp, N., Illig, T., Vogelberg, C., Weiland, S.K., *et al.* (2008). IRF-1 gene variations influence IgE regulation and atopy. *Am J Respir Crit Care Med* 177, 613-621.
- Seki, E., and Brenner, D.A. (2008). Toll-like receptors and adaptor molecules in liver disease: update. *Hepatology* 48, 322-335.

- Serre, D., Gurd, S., Ge, B., Sladek, R., Sinnett, D., Harmsen, E., Bibikova, M., Chudin, E., Barker, D.L., Dickinson, T., *et al.* (2008). Differential allelic expression in the human genome: a robust approach to identify genetic and epigenetic cis-acting mechanisms regulating gene expression. *PLoS Genet* 4, e1000006.
- Southam, L., Rodriguez-Lopez, J., Wilkins, J.M., Pombo-Suarez, M., Snelling, S., Gomez-Reino, J.J., Chapman, K., Gonzalez, A., and Loughlin, J. (2007). An SNP in the 5'-UTR of GDF5 is associated with osteoarthritis susceptibility in Europeans and with in vivo differences in allelic expression in articular cartilage. *Hum Mol Genet* 16, 2226-2232.
- Szabo, G., Dolganiuc, A., and Mandrekar, P. (2006). Pattern recognition receptors: a contemporary view on liver diseases. *Hepatology* 44, 287-298.
- Takeda, K., and Akira, S. (2005). Toll-like receptors in innate immunity. *Int Immunol* 17, 1-14.
- Tamura, T., Yanai, H., Savitsky, D., and Taniguchi, T. (2008). The IRF family transcription factors in immunity and oncogenesis. *Annu Rev Immunol* 26, 535-584.
- Taniguchi, T., Ogasawara, K., Takaoka, A., and Tanaka, N. (2001). IRF family of transcription factors as regulators of host defense. *Annu Rev Immunol* 19, 623-655.
- Vincze, T., Posfai, J., and Roberts, R.J. (2003). NEBcutter: A program to cleave DNA with restriction enzymes. *Nucleic Acids Res* 31, 3688-3691.
- Virta, M., Pessi, T., Helminen, M., Seiskari, T., Kondrashova, A., Knip, M., Hyoty, H., and Hurme, M. (2008). Interaction between CD14-159C>T polymorphism and *Helicobacter pylori* is associated with serum total immunoglobulin E. *Clin Exp Allergy* 38, 1929-1934.
- Von Hahn, T., Halangk, J., Witt, H., Neumann, K., Muller, T., Puhl, G., Neuhaus, P., Nickel, R., Beuers, U., Wiedenmann, B., *et al.* (2008). Relevance of endotoxin receptor CD14 and TLR4 gene variants in chronic liver disease. *Scand J Gastroenterol* 43, 584-592.
- Wietzke-Braun, P., Maouzi, A.B., Manhardt, L.B., Bickeboller, H., Ramadori, G., and Mihm, S. (2006). Interferon regulatory factor-1 promoter polymorphism and the outcome of hepatitis C virus infection. *Eur J Gastroenterol Hepatol* 18, 991-997.
- Yan, H., Yuan, W., Velculescu, V.E., Vogelstein, B., and Kinzler, K.W. (2002). Allelic variation in human gene expression. *Science* 297, 1143.
- Ye, S., Dhillon, S., Ke, X., Collins, A.R., and Day, I.N. (2001). An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Res* 29, E88-88.

Für die fruchtbare Zusammenarbeit und Unterstützung im Zusammenhang mit dieser Doktorarbeit möchte ich mich bei folgenden Personen bedanken:

Zuerst gilt mein besonderer Dank Frau Prof. Dr. rer. nat. Sabine Mihm für die individuelle Anleitung dieser Dissertation und die ständige Diskussionsbereitschaft. Mit konstruktiver Kritik, ihrem Wissen, Hilfe und ihrer Erfahrung hat Sie maßgeblich zum Gelingen dieser Doktorarbeit beizutragen, auch über mehrere Kontinente hinweg. Deshalb meinen herzlichen Dank an Sie!

Meinen Betreuern, Prof. Dr. med. Hunsmann und PD Dr. rer. nat. Michael Hoppert, möchte ich hiermit ein herzliches Dankeschön sagen! Auch durch ihre Hilfsbereitschaft und Motivation war es mir möglich meine Doktorarbeit in dieser Form zu vollenden.

Auch bei Prof. Dr. med. Dr. h. c. Giuliano Ramadori bedanke ich mich nicht nur für die Möglichkeit meine Dissertation in seiner Abteilung durchzuführen, sondern auch für die Bereitschaft mir sein Blut für meine Experimente zu spenden. Darüber hinaus auch an dieser Stelle vielen herzlichen Dank für Ihre persönliche Unterstützung und Ihr Verständnis.

In diesem Zusammenhang danke ich natürlich auch allen bereitwilligen Blutspendern für ihre Unerschrockenheit Nadeln gegenüber.

Mein Dank gilt auch Herrn Dr. Klaus Jung für die Hilfe bei statistischen Fragen und Frau Christine Crozier für die persönliche sprachliche Überarbeitung dieser Dissertation.

Frau Waltraut Kopp danke ich für die Hilfe in technischen Fragen. Besonders aber danke ich an dieser Stelle Corinna Dunaiski, Dr. rer. nat. Martina Blaschke und meiner Mitstreiterin Gesa Martius für ihre Hilfe und Freundschaft auch in anstrengenden Zeiten, sowie allen ungenannten Kollegen und Freunden für die gemeinsame Zeit.

Aber ganz besonders möchte ich meiner Mutter für ihre uneingeschränkte Liebe, Sorge und Unterstützung in allen Bereichen meines Lebens danken. Danke Mam für Alles!

Curriculum vitae

Personal Data

Name	Jasmin Mertens
Date of birth	October, 16th 1978
Place of birth	Munich, Germany
Gender	Female

Education

2007 (Apr) - present University Medical Center Dept. of Gastroenterology and Endocrinology Georg-August University Goettingen Germany	Ph.D. student in Biology (Dr. rer. nat.) Thesis: Exploring the Functional Relevance of Polymorphisms within the CD14 and IRF-1 Gene for Promoter Activity by Haplotype- Specific Chromatin Immunoprecipitation (HaploChIP)
2006 (Feb) - 07 (Mar) Institute of Genomic and Applied Microbiology Georg-August University Goettingen Germany	Diploma in Biology (Dipl. Biol.) Thesis: Identifizierung und Charakterisierung von Genen aus Metagenomen, die Antibiotika- Resistenz vermitteln
2000 - 2006 Georg-August University Goettingen Germany	Studies of Biology Major: Microbiology Minors: Immunology, Botany
1990 - 2000 Comenius Gymnasium Deggendorf Germany	Diploma from German secondary school qualifying for university admission or matriculation (Abitur)